Министерство науки и высшего образования Российской Федерации
ФГБОУ ВО «Уральский государственный лесотехнический университет»

Институт заочного обучения (ИЗО)
Высшая школа биотехнологий
Направление: 19.04.01. Биотехнология
Профиль подготовки: Технология биологически активных веществ и фармпрепаратов на основе растительного сырья

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА

вид работы: магистерская диссертация
на тему:

Использование биоинформационного моделирования при разработке технологий редактирования нуклеотидной последовательности

Выпускник Захаров А.М.
(фамилия, инициалы) (подпись)

Руководитель Тихонов С.Л.
(фамилия, инициалы) (подпись)

Зав. кафедрой Тихонов С.Л.
(фамилия, инициалы) (подпись)

Екатеринбург
2026

Министерство науки и высшего образования Российской Федерации
ФГБОУ ВО «Уральский государственный лесотехнический университет»

Институт заочного обучения (ИЗО)
Высшая школа биотехнологий
Направление: 19.04.01. Биотехнология
Профиль подготовки: Технология биологически активных веществ и фармпрепаратов на основе растительного сырья

ТЕМА ВЫПУСКНОЙ КВАЛИФИКАЦИОННОЙ РАБОТЫ

Использование биоинформационного моделирования при разработке технологий редактирования нуклеотидной последовательности

Утверждена приказом по университету № 2103-ст от 06.10.2025 г.
Кафедра Высшая школа биотехнологии Зав. кафедрой Тихонов С.Л.
Руководитель проекта Тихонов С.Л.
Консультант Тихонов С.Л.
Рецензент Шихалев С.В.
Работа начата «02» июня 2025 г. закончена «11» февраля 2026 г.
Решением кафедры ВШБ от 11 февраля 2026 г. протокол №6 выпускник допущен к защите выпускной квалификационной работы

Зав. кафедрой _____ / С.Л. Тихонов /
«11» февраля 2026 г.

РЕШЕНИЕ
ЭКЗАМЕНАЦИОННОЙ КОМИССИИ

1. Признать, что выпускник Захаров А.М. выполнил(а) и защитил(а) выпускную квалификационную работу с оценкой хорошо

Председатель ЭК _____ / С.В. Чернышева /

Секретарь ЭК _____ / Т.Г. Панова /

Министерство науки и высшего образования Российской Федерации
ФГБОУ ВО «Уральский государственный лесотехнический университет»

Институт заочного обучения (ИЗО)
Высшая школа биотехнологий
Направление: 19.04.01. Биотехнология
Профиль подготовки: Технология биологически активных веществ и фармпрепаратов на основе растительного сырья

УТВЕРЖДАЮ:
Зав. кафедрой Тихонов С.Л.

(подпись)

«2» июня 2025 г.

ЗАДАНИЕ

на выпускную квалификационную работу
Захарова Артура Маратовича
(фамилия, имя, отчество)

1. Вид работы: магистерская диссертация
2. Тема работы: «Использование биоинформационного моделирования при разработке технологий редактирования нуклеотидной последовательности»
утверждена приказом ректора №2103-ст от «06» октября 2025 г.
3. Срок сдачи выпускником законченной работы: «11» февраля 2026 г.
4. Исходные данные: биоинформационное моделирование, нуклеотидная последовательность
5. Содержание расчетно-пояснительной записки:
   - 5.1 Структура организма и объекта исследования
   - 5.2 Специфика уровня нуклеотидной последовательности
   - 5.3 Существующие решения и подходы
   - 5.4 Стратегия кумулятивных децимодификаций
   - 5.5 Биоинформационное моделирование
   - 5.6 Применение в рамках проекта

6. Консультанты по работе, с указанием относящихся к ним разделов:

Раздел	ФИО консультанта	Подпись, дата	
		Задание выдал	Задание принял
Структура организма и объекта исследования	Тихонов С.Л.	02.06.2025	02.06.2025
Специфика уровня нуклеотидной последовательности	Тихонов С.Л.	02.06.2025	02.06.2025
Существующие решения и подходы	Тихонов С.Л.	02.06.2025	02.06.2025
Стратегия кумулятивных децимодификаций	Тихонов С.Л.	02.06.2025	02.06.2025
Биоинформационное моделирование	Тихонов С.Л.	02.06.2025	02.06.2025
Применение в рамках проекта	Тихонов С.Л.	02.06.2025	02.06.2025

7. Календарный план:

№ п/п	Наименование этапов работы	Срок выполнения этапов работы	Примечание
1	Знакомство с проблемой	02.06.2025	
2	Сбор данных, их оценка	02.08.2025	
3	Составление гипотез	02.10.2025	
4	Подготовка выводов	02.12.2025	
5	Оформление ВКР	02.01.2026	
6	Оформление презентации	12.02.2026	
7	Предзащита	13.02.2026	

8. Дата выдачи задания: «2» июня 2025 г.

Руководитель выпускной квалификационной работы

_____ Тихонов С.Л.

(подпись) (фамилия, инициалы)

Задание принял к исполнению

_____ Захаров А.М.

(подпись) (фамилия, инициалы)

6. Консультанты по работе, с указанием относящихся к ним разделов:

Раздел	ФИО консультанта	Подпись, дата	
		Задание выдал	Задание принял
Структура организма и объекта исследования	Тихонов С.Л.	02.06.2025	02.06.2025
Специфика уровня нуклеотидной последовательности	Тихонов С.Л.	02.06.2025	02.06.2025
Существующие решения и подходы	Тихонов С.Л.	02.06.2025	02.06.2025
Стратегия кумулятивных децимодификаций	Тихонов С.Л.	02.06.2025	02.06.2025
Биоинформационное моделирование	Тихонов С.Л.	02.06.2025	02.06.2025
Применение в рамках проекта	Тихонов С.Л.	02.06.2025	02.06.2025

7. Календарный план:

№ п/п	Наименование этапов работы	Срок выполнения этапов работы	Примечание
1	Знакомство с проблемой	02.06.2025	
2	Сбор данных, их оценка	02.08.2025	
3	Составление гипотез	02.10.2025	
4	Подготовка выводов	02.12.2025	
5	Оформление ВКР	02.01.2026	
6	Оформление презентации	12.02.2026	
7	Предзащита	13.02.2026	

8. Дата выдачи задания: «2» июня 2025 г.

Руководитель выпускной квалификационной работы

_____ Тихонов С.Л.

(подпись) (фамилия, инициалы)

Задание принял к исполнению

_____ Захаров А.М.

(подпись) (фамилия, инициалы)

РЕФЕРАТ

Захаров А.М., Использование биоинформационного моделирования при разработке технологий редактирования нуклеотидной последовательности. Магистерская диссертация. ФГБОУ ВО «УГЛТУ», Высшая школа биотехнологий. Руководитель – Тихонов С.Л. – Екатеринбург, 2026.

БИОИНФОРМАЦИОННОЕ МОДЕЛИРОВАНИЕ, РЕДАКТИРОВАНИЕ ГЕНОМА, НУКЛЕОТИДНАЯ ПОСЛЕДОВАТЕЛЬНОСТЬ, СТРАТЕГИЯ КУМУЛЯТИВНЫХ ДЕЦИМОДИФИКАЦИЙ, ГЕННАЯ ТЕРАПИЯ, ВИРУС ЗИКА (ZIKV), МУТАГЕНЕЗ, *IN SILICO* , ИТЕРАТИВНОЕ РЕДАКТИРОВАНИЕ

Данная работа содержит 45 страниц, 17 рисунков, 55 источников, 1 приложение и посвящена новому подходу к редактированию генома. Целью являлось обоснование стратегии кумулятивных децимодификаций – итеративного внесения малых изменений, позволяющего преодолевать защитные механизмы клетки и достигать функционального эффекта. Методология включает системный анализ биологической иерархии и компьютерное моделирование на примере штаммов вируса Зика (ZIKV). Результаты и их новизна заключаются в формализации стратегии и демонстрации её применимости для создания безопасных вирусных векторов на основе модифицированных фрагментов генома. Область применения находится в экспериментальной генетике, генной терапии и биотехнологии.

СОДЕРЖАНИЕ

Обозначения, термины и сокращения 7
Методы исследования 9
Введение 11
1 Структура организма и объекта исследования 12
2 Специфика уровня нуклеотидной последовательности 18
2.1 Наследственные заболевания 18
2.2 Приобретенные заболевания 19
3 Существующие решения и подходы 22
3.1 Пренатальная и постнатальная терапия 22
3.2 Инвазивные и неинвазивные методы 24
3.3 Варианты и время применения вакцин 26
4 СТРАТЕГИЯ КУМУЛЯТИВНЫХ ДЕЦИМОДИФИКАЦИЙ 28
5 Биоинформационное моделирование 31
5.1 Вирус Зика (ZIKV) 31
5.2 Анализ последовательности и выбор области редактирования 35
5.3 Построение вирусного вектора 40
6 Применение в рамках проекта 41
Результаты и выводы 42
Список использованных источников и литературы 43
Приложение А 47

ОБОЗНАЧЕНИЯ, ТЕРМИНЫ И СОКРАЩЕНИЯ

Апоптоз – программируемая гибель клетки, запускаемая при необратимых повреждениях ДНК или нарушении клеточного гомеостаза; механизм защиты организма от потенциально опасных мутантных клеток.

Вирусный вектор – генетическая конструкция на основе вируса, используемая для доставки терапевтических нуклеиновых кислот или редактирующих компонентов в клетки.

Ген – участок нуклеотидной последовательности ДНК (или РНК у некоторых вирусов), кодирующий структурно и функционально обособленную молекулу, чаще белок или РНК.

Геном – совокупность всей наследственной информации организма, записанной в виде последовательности нуклеотидов ДНК (или РНК у РНК-содержащих вирусов).

Генная терапия – совокупность биотехнологических методов, направленных на изменение генетической программы клеток для лечения наследственных, онкологических или инфекционных заболеваний.

Делеция – тип мутации, при которой происходит утрата (выпадение) одного или нескольких нуклеотидов из нуклеотидной последовательности.

Децимодификация – авторский термин, обозначающий изменение не более чем одной десятой части (10%) от выбранного целевого участка нуклеотидной последовательности за один цикл редактирования.

Инсерция – тип мутации, при которой в нуклеотидную последовательность встраиваются один или несколько дополнительных нуклеотидов.

Метилирование ДНК – эпигенетическая модификация, заключающаяся в присоединении метильной группы к цитозину (обычно в составе CpG-динуклеотидов); часто служит сигналом для подавления транскрипции гена.

Миссенс-мутация – точечная мутация, приводящая к замене одной аминокислоты на другую в кодируемом белке.

Нонсенс-мутация – точечная мутация, при которой кодон, кодирующий аминокислоту, превращается в стоп-кодон, что приводит к преждевременному обрыву синтеза белка.

Нуклеотидная последовательность – линейная последовательность мономеров (нуклеотидов A, T, G, C для ДНК; A, U, G, C для РНК), представляющая собой первичную структуру молекулы ДНК или РНК и кодирующая генетическую информацию.

Репликация – процесс удвоения молекулы ДНК (или РНК), обеспечивающий передачу генетической информации дочерним клеткам.

Репарация ДНК – совокупность внутриклеточных механизмов, обеспечивающих обнаружение и исправление повреждений в молекуле ДНК.

Сдвиг рамки считывания (фреймшифт) – нарушение триплетности чтения кода, вызванное инсерцией или делецией нуклеотидов, число которых не кратно трём.

Стратегия кумулятивных децимодификаций – форма редактирования нуклеотидной последовательности, при которой каждое множественное вмешательство ограничено изменением десятой части от целевой области, а их кумулятивный эффект достигается за счет преодоления барьеров и итеративности.

Технология – совокупность методов и инструментов для достижения желаемого результата, способ преобразования данного в необходимое.

Транскрипция – процесс синтеза молекулы РНК на матрице ДНК, первый этап реализации генетической информации.

Трансляция – процесс синтеза полипептидной цепи (белка) на рибосоме с использованием матричной РНК (мРНК) в качестве шаблона.

Хроматин – нуклеопротеидный комплекс, состоящий из ДНК, гистонов и негистоновых белков; структурная основа хромосом, определяющая доступность генетического материала для транскрипции и репарации.

Эпигенетика – раздел биологии, изучающий наследуемые изменения активности генов, не связанные с изменением первичной нуклеотидной последовательности ДНК/

Ex vivo – подход в генной терапии, при котором клетки пациента извлекаются из организма, подвергаются модификации в лабораторных условиях и затем возвращаются обратно.

In silico – исследование биологических процессов с помощью компьютерного моделирования и биоинформатических методов.

In vivo – подход в генной терапии, при котором генетический материал доставляется непосредственно в организм пациента.

МЕТОДЫ ИССЛЕДОВАНИЯ

Данная работа выполнена в рамках теоретико-методологического исследования с использованием методов *in silico*. Методологическая основа работы базируется на системном анализе, сравнительной геномике и математическом моделировании. Исследование проведено в несколько этапов.

1. Общенаучные и теоретические методы

На первом этапе работы для анализа современного состояния проблемы редактирования генома и классификации существующих подходов использовались методы теоретического анализа и синтеза. Был проведён критический обзор научной литературы, посвященной пренатальной и постнатальной терапии, инвазивным и неинвазивным методам доставки, а также молекулярным механизмам наследственных и приобретенных заболеваний. Метод системного анализа позволил рассмотреть человеческий организм как иерархическую метасистему (от уровня нуклеотидной последовательности до целостного организма), что необходимо для понимания системных последствий точечных генетических вмешательств. На основе синтеза данных о защитных механизмах клетки и существующих ограничениях геномного редактирования была сформулирована гипотеза о возможности применения стратегии кумулятивных децимодификаций.

2. Методы биоинформационного моделирования

Основным инструментом проверки гипотезы стало биоинформационное моделирование, выполненное с использованием открытых баз данных и специализированного программного обеспечения.

Объектом моделирования выбран вирус Зика (ZIKV) ввиду его медико-биологической значимости, компактного генома (представленного одноцепочечной РНК) и наличия полногеномных данных в открытых базах. Нуклеотидные последовательности девяти штаммов вируса Зика (включая референтные штаммы MR766 и азиатские изоляты, вызвавшие пандемию 2015–2016 гг.) были получены из базы данных GenBank NCBI (номера доступа: LC369584.1, LC219720.1, LC191864.1, LC190723.1, LC002520.1, MK241417.1, MK241416.1, MK241415.1, NC_012532.1).

Для выявления консервативных и вариабельных участков генома, а также для оценки генотипической вариабельности штаммов было проведено множественное выравнивание нуклеотидных последовательностей. Выравнивание выполнялось с использованием алгоритма Clustal Omega с последующей визуальной коррекцией. На

основе выравнивания был проведён анализ нуклеотидного состава (GC-состав), поиск однонуклеотидных полиморфизмов (SNP) и потенциальных сайтов, ассоциированных с патогенностью.

Для оценки потенциального влияния нуклеотидных замен на функцию белков проводилось предсказание вторичной структуры РНК и влияния миссенс-мутаций на структуру белков. Использовались инструменты предсказания вторичной структуры РНК (например, алгоритм минимальной свободной энергии в программе RNAfold) и гомологичное моделирование белков (SWISS-MODEL) для оценки расположения ключевых аминокислотных остатков в каталитических доменах вирусных полимераз.

3. Моделирование стратегии кумулятивных децимодификаций

На основе правила Парето и принципа итеративности была предложена математическая модель, описывающая процесс последовательных изменений генома.

Для описания нелинейного прироста функциональной эффективности по мере накопления малых изменений была предложена формула (4.1), учитывающая контекстный коэффициент адаптации ($C\sim n\sim$). Математическое моделирование позволяет визуализировать разницу между линейной суммой изменений (арифметической) и потенциальным кумулятивным эффектом (геометрической прогрессией), возникающим за счет подготовки контекста последовательности к последующим заменам.

4. Моделирование применения

На заключительном этапе, в рамках раздела «Применение в рамках проекта», на основе мысленного эксперимента был смоделирован процесс создания гипотетического вирусного вектора, несущего запрограммированные изменения. Это позволяет оценить стадию применимости стратегии.

ВВЕДЕНИЕ

Человек несовершенен и, как всякая сложная система, имеет структурные недостатки, проявляющиеся как в наследственных, так и приобретенных заболеваниях – от моногенных патологий до онкологических процессов. Стремление исправить их породило направление генной терапии и технологию редактирования генома, вершиной которой сейчас являются системы на основе CRISPR-Cas.

Однако существующие технологии сталкиваются с фундаментальными ограничениями. Высокая эффективность нуклеаз часто сопровождается низкой специфичностью и риском внецелевых повреждений ДНК. Кроме того, клеточные системы репарации распознают разрывы как критическое повреждение, что запускает сложные и зачастую непредсказуемые механизмы восстановления, приводящие к нежелательным делециям, инсерциям или хромосомным перестройкам.

В связи с этим возникает гипотеза: можно ли добиться направленного изменения функции гена, не создавая критических реакций или даже используя их в нужном направлении?

1 СТРУКТУРА ОРГАНИЗМА И ОБЪЕКТА ИССЛЕДОВАНИЯ

Человеческий организм на макромасштабе (см. рисунок 1.1) представляет собой интегрированную и саморегулирующуюся биологическую систему высочайшей сложности – метасистему [1]. Это **уровень фенотипа** – совокупности всех наблюдаемых и измеряемых характеристик, формирующихся в результате динамического взаимодействия генотипа с окружающей средой. Здесь организм предстает не как совокупность органов, а как единое целое, обладающее эмерджентными свойствами – сознание, иммунный ответ, системная регуляция гомеостаза, – которые нельзя свести к функциям отдельных его частей [2].

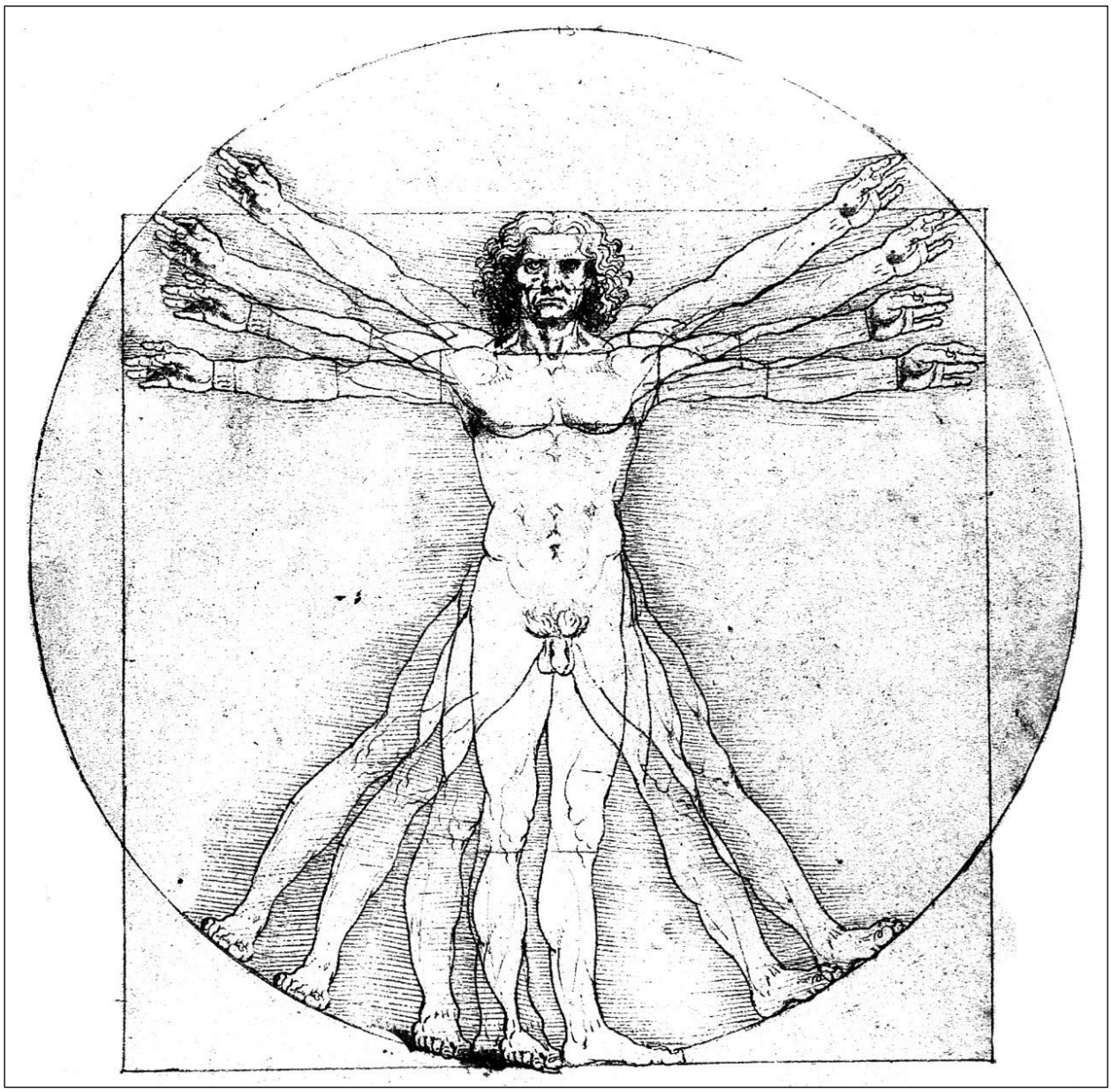

Рисунок 1.1 – Человек Леонардо да Винчи

На этом системном уровне проявляются те самые «структурные недостатки», упомянутые во введении. Наследственные моногенные заболевания (например, муковисцидоз, серповидноклеточная анемия), сложные мультифакториальные патологии (аутоиммунные процессы, сердечно-сосудистые заболевания) и онкологические процессы являются, по сути, системными сбоями. Эти сбои возникают из-за дисфункции на нижележащих уровнях, но их клинические проявления – будь то дыхательная недостаточность, хронический болевой синдром или кахексия – есть результат нарушения слаженной работы множества взаимосвязанных подсистем.

Нисходя от целостного организма, мы сталкиваемся с уровнем его структурной организации – **уровнем органов и тканей**. Это уровень функциональной специализации и морфологического разнообразия (см. рисунок 1.2).

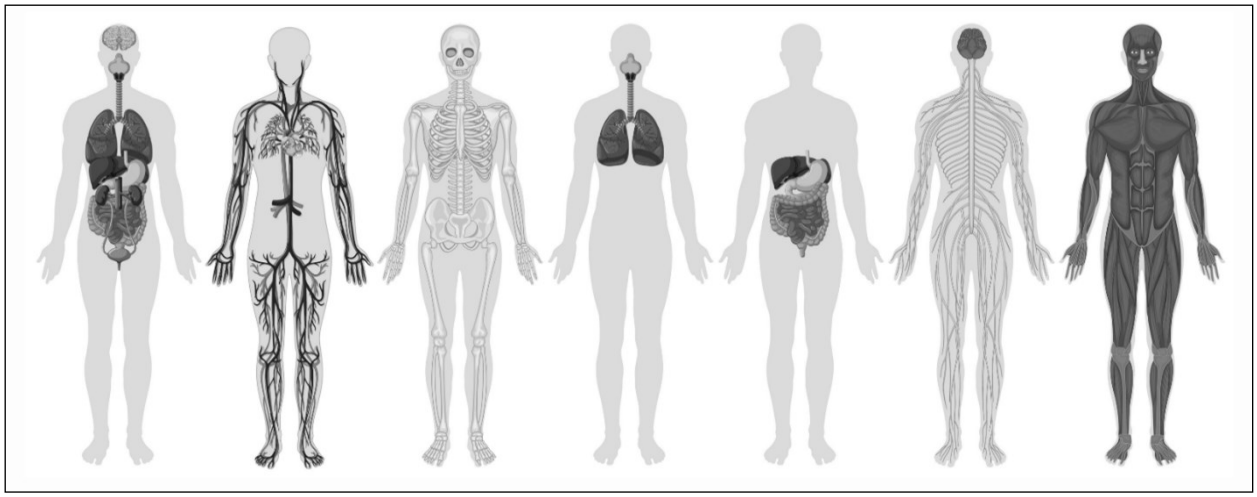

Рисунок 1.2 – Органы и системы

Если организм как система задает общую цель, то органы – ее конкретные исполнительные модули: сердце обеспечивает гемодинамику, печень – метаболический гомеостаз, нервная ткань – проведение сигнала. Каждый орган представляет собой сложный ансамбль из нескольких типов тканей (эпителиальной, соединительной, мышечной, нервной), находящихся в строго упорядоченных пространственных взаимоотношениях [3].

Именно на этом гистологическом уровне наследственная или приобретенная патология обретает свою конкретную морфологию – гипертрофия миокарда, фиброз печени, амилоидные отложения, злокачественная опухоль.

Спускаясь по иерархии биологической организации, мы достигаем фундаментального уровня, на котором генетическая информация обретает функциональность, – **клеточного уровня**. Клетка представляет собой минимальную автономную единицу жизни, способную к самоподдержанию, репликации и реагированию на стимулы (см. рисунок 1.3). Это сложно

организованный гетерогенный компартмент, где ламинарный поток генетической информации сталкивается с турбулентностью метаболизма и жесткими ограничениями физико-химических законов. В контексте редактирования генома клетка является не пассивной мишенью, а активным, высокореактивным участником процесса, чьи внутренние системы во многом определяют успех или неудачу любого вмешательства [4].

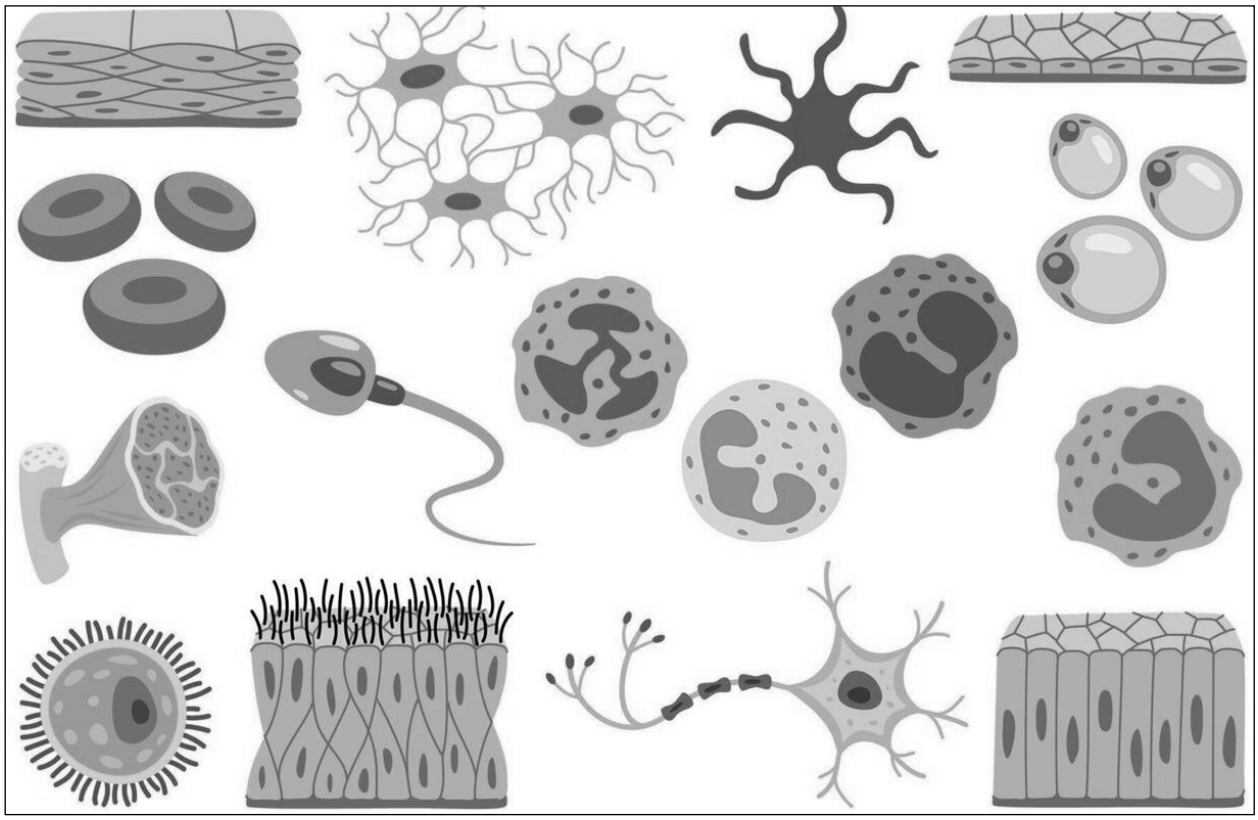

Рисунок 1.3 – Разнообразие клеток человека

На пути от линейной генетической инструкции к ее функциональной реализации молекула ДНК претерпевает радикальную трансформацию. Она не существует в клеточном ядре в виде вытянутой нити, но организована в сложнейшую иерархическую структуру – **хроматин** (см. рисунок 1.4). Этот уровень представляет собой принципиальный интерфейс между химией нуклеотидной последовательности и биологией ее считывания. Хроматин – это динамическая «упаковка» генома, определяющая фундаментальный параметр: доступность генетической информации для систем репарации, транскрипции и, что критически для нашей задачи, для инструментов редактирования [5].

На **уровне взаимодействий макромолекул** живая система предстает не как анатомическая или клеточная структура, а как динамическая сеть, управляемая законами стерического соответствия и химической аффинности (см. рисунок 1.5). Это уровень, на котором информация, закодированная в нуклеотидной последовательности, материализуется в

конкретные биохимические функции. Центральная догма молекулярной биологии (ДНК → РНК → белок) задает основное направление потока, но реальная картина бесконечно сложнее: это переплетение транскрипции, сплайсинга, трансляции, посттрансляционных модификаций, а также регуляции на каждом из этих этапов [6].

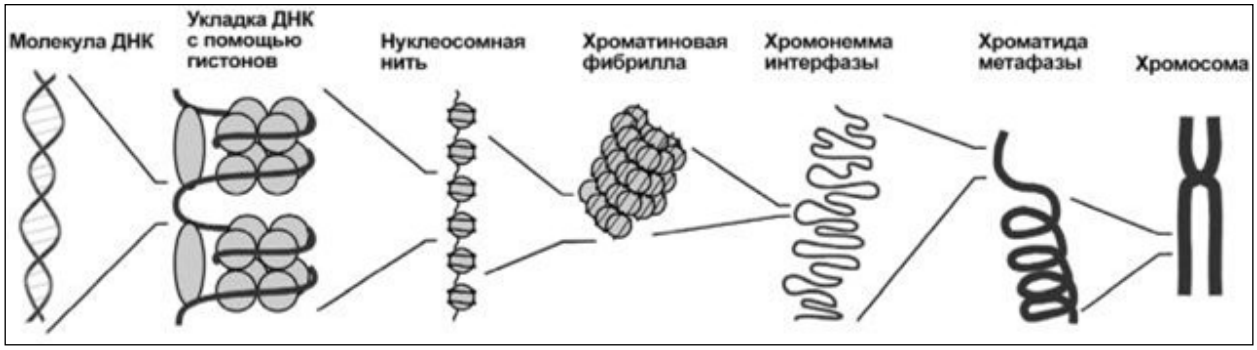

Рисунок 1.4 – Этапы организации хроматина

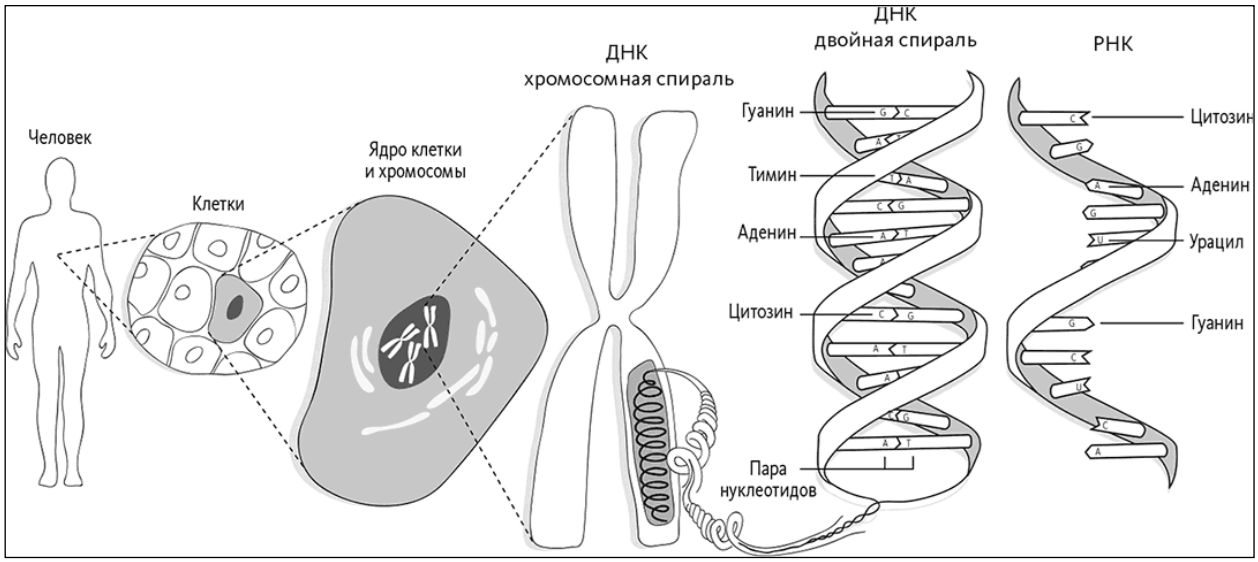

Рисунок 1.5 – От организма к молекулам

На самом глубоком уровне разрешения биологическая сложность сводится к дискретной, линейной записи информации (см. рисунок 1.6). **Геном человека** – это последовательность длиной приблизительно 3,2 миллиарда пар химических знаков, четырехбуквенный алфавит которого (аденин [A], тимин [T], гуанин [G], цитозин [C]) кодирует всю инструкцию для развития и функционирования организма. Это уровень первичной генетической детерминации, где наследуемые «ошибки» – однонуклеотидные полиморфизмы, делеции, инсерции и мутации – получают свое материальное воплощение. Именно здесь, в этой текстовой строке, и разворачивается основное действие любой технологии редактирования генома [7].

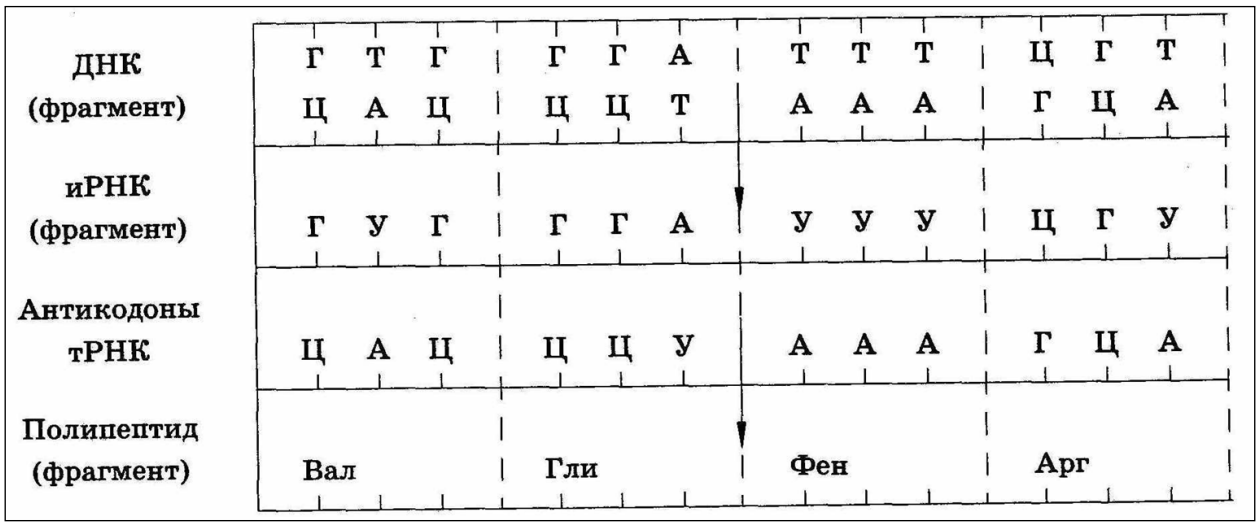
ДНК (фрагмент)
Г Т Г Г Г А Т Т Т Ц Г Т
Ц А Ц Ц Ц Т А А А Г Ц А
иРНК (фрагмент)
Г У Г Г Г А У У У Ц Г У
Антикодоны тРНК
Ц А Ц Ц Ц У А А А Г Ц А
Полипептид (фрагмент)
Вал
Гли
Фен
Арг

Рисунок 1.6 – Запись нуклеотидной последовательности

Для лучшего понимания приведём следующую аналогию: Представьте, что каждая наша клетка – это огромная, очень сложная фабрика. Для своей работы ей нужны чертежи и инструкции. Эти инструкции – наш геном.

Геном – это полная инструкция по созданию и функционированию организма, записанная на молекуле ДНК.

ДНК – это длинная, скрученная в спираль молекула. Ее «буквами» служат четыре химических основания: Аденин (A), Тимин (T), Гуанин (G) и Цитозин (C). Последовательность этих букв – нуклеотидная последовательность – и есть генетический код. А всегда соединяется с T, а G с C, образуя «ступеньки» знаменитой двойной спирали. Эта точная парность – ключ к копированию и чтению информации.

Геном – это не сплошной текст. Он разбит на главы и параграфы – гены. Ген – это конкретный отрезок ДНК, несущий инструкцию для создания определенной молекулы, чаще всего белка. Белки – это «рабочие лошадки» клетки: они строят структуры, ускоряют реакции (ферменты), передают сигналы.

Как читается инструкция? Процессы транскрипции и трансляции. Чтобы использовать инструкцию, клетка не трогает ценную оригинальную ДНК в ядре. Вместо этого она ее «переписывает»:

– Транскрипция. Специальный фермент «прочитывает» ген и создает его временную копию – молекулу матричной РНК (мРНК). Это как сделать фотокопию одной страницы из книги и отнести ее в цех.

– Трансляция. На «фабричном цехе» клетки – рибосоме – эта копия (мРНК) используется как шаблон для сборки белков из аминокислот. Каждые три буквы РНК (кодон) кодируют одну конкретную аминокислоту.

Клетке не нужны все белки одновременно. Клетке кожи не нужны белки, отвечающие за зрение. Поэтому гены регулируются – их можно включать и выключать. За это отвечают специальные участки ДНК (промоторы, энхансеры) и транскрипционные факторы – белки, которые садятся на эти участки и либо разрешают, либо запрещают процесс транскрипции.

Представьте, что текст в книге (последовательность ДНК) остается неизменным, но некоторые параграфы заклеены стикерами или выделены маркером. Это – эпигенетика. Это химические модификации, которые не меняют последовательность букв A, T, G, C, но влияют на то, насколько активно читается ген:

– Метилирование ДНК. Чаще всего это добавление метильной группы (-CH3) к основанию Цитозин (C). Обычно это «стоп-сигнал», который мешает считыванию гена и выключает его.

– Модификации гистонов. ДНК не плавает в ядре свободно, она намотана на белковые «катушки» – гистоны. Эти катушки можно химически модифицировать (например, ацетилировать). Если модификации делают намотку более рыхлой – ген легче прочитать (включен). Если намотка становится плотнее – ген закрыт и не работает.

Структура ДНК и белки, которые с ней связываются

ДНК – не просто линейная нить. Она образует сложные трехмерные петли и структуры, которые подводят далекие гены или их регуляторы близко друг к другу. Это влияет на их работу. Многие белки (как те же транскрипционные факторы) должны физически связаться с определенной короткой последовательностью ДНК (например, «AGGTCA»), чтобы выполнить свою функцию. Изменение даже одной буквы в этом сайте связывания может помешать белку «пристыковаться» и тем самым нарушить регуляцию гена.

Геном – главная ценность. Клетка охраняет его двумя основными способами:

– Системы репарации (починки) ДНК. Они постоянно сканируют ДНК на наличие ошибок и повреждений. Если найден разрыв одной или двух цепей ДНК, система репарации пытается его «зашить», иногда с ошибками или используя в качестве шаблона похожую хромосому.

– Апоптоз (программируемая гибель клетки). Если повреждения ДНК слишком серьезны и их не исправить, клетка получает сигнал на самоуничтожение. Это защищает организм от потенциально опасных мутаций, которые могут привести, например, к раку.

Любая технология направленного воздействия на геном – это попытка точечно изменить один из описанных выше элементов:

– Последовательность ДНК (заменить буквы).

– Эпигенетические пометки (добавить или убрать «стикеры»).

– Доступность гена (через модификацию гистонов или связывающихся белков).

После данной аналогии имеет смысл перечитать работу с начала.

2 СПЕЦИФИКА УРОВНЯ НУКЛЕОТИДНОЙ ПОСЛЕДОВАТЕЛЬНОСТИ

Выше мы уже коснулись темы редактирования генома. Однако это неизбежно подводит нас к более глубокому и фундаментальному вопросу: а зачем, собственно, человеку понадобилось вторгаться в собственный генетический код? Что мы пытаемся в нем исправить? Важно понять не только причины несовершенства нашей ДНК, но и сам механизм возникновения этих «поломок» – как тех, что происходят внезапно, так и тех, что накапливаются разрушительной динамикой на протяжении жизни (см. рисунок 2.1-2.2).

2.1 Наследственные заболевания

В основе наследственных заболеваний лежат стойкие изменения нуклеотидной последовательности ДНК – мутации, которые передаются через половые клетки. Эти изменения можно классифицировать по их структурной сути. Наиболее частыми являются точечные мутации, когда один нуклеотид заменяется на другой [8]. Последствия зависят от локализации. Если она происходит в кодирующей части гена, возможны три сценария.

Во-первых, синонимическая мутация, при которой из-за вырожденности генетического кодона изменённый триплет продолжает кодировать ту же аминокислоту, и функция белка не нарушается.

Во-вторых, миссенс-мутация, когда изменение приводит к кодированию другой аминокислоты в полипептидной цепи. Эффект варьируется от минимального, если свойства аминокислоты схожи, до катастрофического, если замена критически меняет трёхмерную структуру белка. Классический пример – серповидноклеточная анемия, где замена всего одного нуклеотида (аденина на тимин) в гене бета-глобина приводит к подстановке валина вместо глутаминовой кислоты, в результате чего гемоглобин полимеризуется и эритроциты деформируются [9].

В-третьих, нонсенс-мутация – самая грубая, при которой нормальный кодон превращается в стоп-кодон. Это приводит к преждевременному обрыву синтеза белка, создавая укороченный и, как правило, не функциональный продукт, что встречается, например, при муковисцидозе [10, 11].

Более серьёзные последствия, чем единичная замена, несут вставки или делеции одного или нескольких нуклеотидов. Если их число не кратно трём, происходит сдвиг рамки считывания. Рибосома начинает читать последовательность с неправильной позиции, и все последующие кодоны, а значит и аминокислоты, изменяются. Это почти

гарантированно приводит к синтезу абсолютно нефункционального белка или появлению нового стоп-кодона. Такой механизм лежит в основе многих тяжёлых заболеваний, например, некоторых форм болезни Тея-Сакса [12].

Отдельный и коварный класс мутаций – это экспансия (расширение) тандемных повторов, когда короткая последовательность нуклеотидов (вроде CAG или CGG) в регуляторной или кодирующей области гена многократно копируется от поколения к поколению. Достигая критического порога, эти повторы нарушают экспрессию гена или приводят к синтезу белка с токсичными свойствами, как это происходит при болезни Хантингтона (повторы CAG в гене HTT) или синдроме ломкой Х-хромосомы [13, 14].

Каким же образом эти молекулярные поломки превращаются в конкретное заболевание? Механизм напрямую связан с типом наследования и функцией повреждённого белка.

При рецессивных заболеваниях, таких как фенилкетонурия или муковисцидоз, мутация обычно приводит к полной или почти полной потере функции белка [15]. Однако для проявления болезни необходимы две копии дефектного гена, так как одной рабочей копии часто достаточно для поддержания нормального фенотипа. Носители с одной мутацией, как правило, здоровы.

В случае доминантного наследования, характерного для болезни Хантингтона или нейрофиброматоза, одной мутантной копии гена достаточно для манифестации патологии. Здесь механизмы иные: часто это либо «доминантно-негативный» эффект, когда продукт мутантного аллеля нарушает работу нормального белка, либо, что ещё опаснее, приобретение белком новой токсической функции, как при образовании агрегатов, убивающих нейроны при хорее Хантингтона [16, 17].

Отдельная картина наблюдается при заболеваниях, сцепленных с Х-хромосомой (гемофилия А, мышечная дистрофия Дюшенна). Поскольку у мужчин только одна Х-хромосома, даже рецессивная мутация в её генах не имеет здоровой пары для компенсации и потому проявляется. У женщин вторая, нормальная Х-хромосома обычно обеспечивает достаточный уровень функционального белка, чтобы избежать тяжёлых симптомов.

Таким образом, путь от конкретной нуклеотидной «опечатки» к клинической картине предопределён как характером самой мутации, так и её положением в геноме и биологической ролью белка-мишени.

2.2 Приобретенные заболевания

В отличие от наследственных болезней, приобретённые заболевания возникают из-за соматических мутаций – изменений в последовательности ДНК, которые происходят в

клетках тела (соматических) на протяжении жизни человека. Эти мутации накапливаются в отдельных клетках-потомках, что делает их ключевым механизмом развития рака, аутоиммунных расстройств и ряда других состояний, а также биологической основой старения. Источником таких мутаций служат неизбежные ошибки при репликации ДНК и, что особенно важно, воздействие факторов внешней среды: ультрафиолетового и ионизирующего излучения, химических канцерогенов (например, из табачного дыма), некоторых вирусов и процессов хронического воспаления [18, 19].

Типы самих мутаций в нуклеотидной последовательности схожи с наследственными, но их контекст и последствия иные. Чаще всего происходят точечные замены, вызванные, например, ошибкой полимеразы или химическим повреждением основания. Ультрафиолетовый свет индуцирует характерную замену цитозина на тимин в динуклеотидах ЦрГ, что является молекулярной «подписью» меланомы и других видов рака кожи [20]. Канцерогены табачного дыма, такие как бензопирен, вызывают специфичные замены гуанина на тимин в определённых генах (например, TP53 и KRAS), что ведёт к неконтролируемому росту клеток лёгких [21, 22]. Не менее опасны вставки и делеции (инделы), приводящие к сдвигу рамки считывания или потере критических участков генов. Кроме того, для приобретённых заболеваний, особенно онкологических, крайне характерны крупные хромосомные перестройки: транслокации, делеции, амплификации и инверсии. Они часто возникают из-за неправильного восстановления двунитевых разрывов ДНК.

Механизм, через который эти случайные мутации превращаются в заболевание, принципиально отличается от менделевского наследования и основан на концепции клональной селекции. В тканях организма постоянно происходят мутации, но если они затрагивают ключевые гены, регулирующие деление, репарацию ДНК и гибель клетки, это даёт такой клетке селективное преимущество для выживания и размножения [26].

Существует несколько основных путей:

– Активация онкогенов через приобретённую доминантную мутацию

Для этого достаточно повреждения одной копии гена в соматической клетке. Мутация (часто точечная замена) превращает нормальный прото-онкоген, регулирующий рост, в постоянно активный онкоген. Например, мутация в гене BRAF (замена валина на глутаминовую кислоту в положении 600) приводит к его постоянной активации и служит драйвером меланомы [23].

– Инактивация генов-супрессоров по приобретённо-рецессивному механизму

Эти гены действуют как «тормоза» клеточного цикла опухолей. Чтобы их выключить, нужно повредить обе копии гена в одной клетке. Первая мутация может быть унаследована

(как при наследственных синдромах) или возникнуть соматически, а вторая – почти всегда приобретённое событие, например, крупная делеция или потеря целой хромосомы. Классический пример – ген TP53, «страж генома», мутации которого обнаруживаются более чем в половине всех видов рака [24].

– Нарушение систем стабильности генома

Приобретённые мутации в генах, отвечающих за репарацию ДНК (например, системе неспаренных оснований – MMR), приводят к состоянию «микросателлитной нестабильности». Клетка теряет способность исправлять ошибки, и в её ДНК с огромной скоростью накапливаются новые мутации, в том числе в онкогенах и супрессорах, что резко ускоряет развитие рака [25].

Таким образом, приобретённые заболевания – это следствие накопления соматических мутаций в ДНК отдельных клеток под давлением внешних факторов и внутренних ошибок. Эти мутации, нарушающие тонкий баланс между делением, дифференцировкой и гибелью клеток, дают им селективное преимущество, приводя к формированию патологического клона, который и становится основой опухоли или дисфункции ткани. Процесс носит многоступенчатый характер и требует накопления нескольких «драйверных» мутаций в ключевых генах, что объясняет, почему риск многих приобретённых заболеваний резко возрастает с возрастом [26].

3 СУЩЕСТВУЮЩИЕ РЕШЕНИЯ И ПОДХОДЫ

Сама логика вещей подсказывает: если в системе обнаруживаются сбои, их необходимо устранять. Однако, когда речь заходит о критическом накоплении таких «сбоев» в цепочке ДНК, а естественные механизмы организма не справляются, возникает более смелый и фундаментальный вопрос – что если полностью обновить устаревший «код», заменив его новой, неповрежденной версией?

3.1 Пренатальная и постнатальная терапия

Пренатальная и постнатальная терапия, направленная на коррекцию нуклеотидной последовательности, представляют собой два принципиально разных временных подхода в лечении наследственных и приобретенных заболеваний. Их разделяет ключевое событие – рождение, определяющее не только доступ к организму, но и этические рамки, сложность вмешательства и его потенциальную масштабность [27].

Пренатальная терапия (внутриутробная) ставит целью коррекцию генетической аномалии на самых ранних стадиях развития организма (см. рисунок 3.1). Её преимущество – возможность воздействовать на большую популяцию клеток, включая зародышевые, или на развивающийся орган до появления необратимых патологических изменений.

Рисунок 3.1 – Пренатальная терапия

При наследственных заболеваниях, вызванных известными мутациями, теоретически возможно применение методов редактирования генома (например, CRISPR-Cas) на стадии эмбриона для исправления дефекта во всех клетках будущего организма, что приведёт к полному излечению и исключит передачу потомству [28]. Такие вмешательства сопряжены с этическими рисками и в большинстве стран мораторированы [29].

Более реалистичный современный подход – соматическая генная терапия in utero, когда терапевтический ген доставляется в специфические ткани плода (например, при болезнях печени или метаболизма), позволяя организму к моменту рождения выработать необходимый белок и избежать манифестации болезни [30].

Для приобретённых заболеваний (например, некоторых вирусных инфекций) пренатальная терапия может заключаться в доставке антивирусных агентов или защитных генов. Основные риски включают иммунный ответ плода, потенциальное воздействие на развивающиеся органы и риск нецелевого редактирования.

Постнатальная терапия начинается после рождения и в настоящее время является основным полем применения генетических и клеточных технологий.

При наследственных заболеваниях она направлена на коррекцию последствий генетического дефекта в соматических клетках. Это достигается путём ex vivo модификации собственных клеток пациента (например, гемопоэтических стволовых клеток при лейкодистрофиях) с последующей реинфузией, либо in vivo доставкой ген-терапевтических конструктов (например, при спинальной мышечной атрофии) [31, 32].

Для приобретённых заболеваний, таких как онкологические, постнатальная терапия на уровне нуклеотидов включает CAR-T-клеточную терапию, где лимфоциты пациента генетически перепрограммируются для атаки опухоли, или редактирование генов для повышения устойчивости к инфекциям (как в случае с ВИЧ) [33, 34]. Главное ограничение постнатального подхода – невозможность исправить все клетки организма, особенно при поражённых органах со слабым клеточным обновлением (мозг, мышцы). Терапия часто направлена на купирование симптомов или замещение функции, а не на полное устранение первопричины, и может требовать повторных введений [35].

Таким образом, пренатальная терапия теоретически предлагает радикальное решение для наследственных болезней через раннее и масштабное вмешательство, но остаётся областью сложных этических дилемм и технологических барьеров. Постнатальная терапия, будучи более безопасной в этическом плане и уже вошедшей в клиническую практику, часто является корректирующей, а не искореняющей, и борется с последствиями устоявшейся патологии.

3.2 Инвазивные и неинвазивные методы

Инвазивные и неинвазивные методы в контексте коррекции нуклеотидной последовательности представляют собой дихотомию, основанную на способе доставки терапевтического агента к клеткам-мишеням и, в пренатальном периоде, на степени физического проникновения в маточно-плодный комплекс. Это разделение определяет не только сложность процедуры и риск осложнений, но и принципиальную достижимость тех или иных тканей [36].

Инвазивные методы подразумевают прямое физическое проникновение в организм или биологическую среду для доставки генетического материала. В пренатальной терапии к ним относятся внутриутробные инъекции: внутрипортальное, внутримышечное или внутричерепное введение вирусных векторов под контролем УЗИ, а также инъекции непосредственно в желточный мешок или пуповину [37]. Эти подходы позволяют преодолеть гистогематические барьеры и доставить конструкт в труднодоступные зоны плода, но сопряжены с риском инфицирования, преждевременных родов и травматизации.

В постнатальной практике классическим инвазивным методом является *ex vivo* генная терапия, требующая хирургического забора (аферез, биопсия) стволовых клеток костного мозга или печени, их генетической модификации в лаборатории и последующей внутривенной реинфузии после миелоаблативной химиотерапии. Инвазивность здесь смещена на этап забора клеток и подготовки организма. Также сюда относят прямые интратекальные или интрацеребральные инъекции векторов для лечения нейродегенеративных заболеваний, требующие люмбальной пункции или нейрохирургического вмешательства.

Неинвазивные методы, в свою очередь, исключают хирургическую травму и прямое проникновение в организм через разрезы или пункции крупных полостей. Основным инструментом здесь выступает системная внутривенная инфузия (см. рисунок 3.2).

В постнатальном периоде это стандарт для большинства одобренных генных терапий: капельное введение аденоассоциированных вирусных векторов периферически или через центральный венозный катетер. Вектор циркулирует в кровотоке, проникает в ткани-мишени (обычно печень, мышцы) через капиллярную сеть.

В пренатальной терапии неинвазивность относительна: она достигается путем внутривенного введения матери, после чего вектор должен преодолеть плацентарный барьер, чтобы достичь плода. Это минимально травматично для плода, но крайне неэффективно из-за фильтрации крови плацентой и иммунного ответа матери. Истинно неинвазивными можно считать пероральные формы генетических конструктов (например,

инактивированные бактерии с плазмидами), однако их биодоступность для системного редактирования нуклеотидных последовательностей пока крайне низка.

Рисунок 3.2 – Неинвазивная терапия

Таким образом, фундаментальное различие между подходами лежит в балансе «доступность тканей» против «риск и системность». Инвазивные методы оправданы там, где необходимо либо воздействовать на изолированные компартменты (мозг, спинномозговая жидкость, костный мозг) либо провести полную замену клеточного генофонда (*ex vivo*), что невозможно сделать системным введением без удержания в печени. Неинвазивная системная доставка, напротив, позволяет одномоментно обработать весь организм, но ограничена тропизмом вектора и невозможностью преодолеть гематоэнцефалический барьер без специализированных капсидов [38].

3.3 Варианты и время применения вакцин

При выборе постнатальной терапии и неинвазивных методов, как наиболее безопасной комбинации, остаётся не так много вариантов для применения. Основными из них являются вакцины [39].

Рисунок 3.3 – Варианты упаковки вакцин

В классической вакцинологии время вмешательства относительно момента инфицирования является основным фактором, определяющим не только эффективность, но и сам механизм защиты. Традиционно выделяют три стратегических окна для применения: профилактическое (до инфицирования), постконтактное (раннее после инфицирования) и терапевтическое (на стадии развившегося заболевания и осложнений) [40].

Наиболее распространенный и эффективный подход – это **профилактическая вакцинация**. Целью является формирование в организме «иммунологической памяти» задолго до встречи с реальным патогеном. Механизм действия основан на введении в организм антигена (ослабленного или инактивированного возбудителя, его субъединиц, или, в новейших технологиях, матричной РНК, кодирующей этот антиген).

Ключевые особенности:

– Создается пул долгоживущих В- и Т-лимфоцитов, специфичных к патогену.

– При заражении иммунная система распознает знакомый антиген и запускает ускоренный синтез защитных антител (IgG), не тратя время на первичное распознавание.

– Массовая профилактическая вакцинация способна создавать коллективный иммунитет, прерывая цепи передачи инфекции (например, корь, полиомиелит).

Когда заражение уже произошло, но клинические симптомы еще не развились, применяется **постконтактная вакцина**. Здесь она работает как инструмент для «ускорения» иммунного ответа, чтобы опередить размножение патогена [41].

Механизм может быть двояким:

– Введение вакцины сразу после контакта стимулирует быструю выработку антител. Это эффективно, если инкубационный период достаточно длителен (как при бешенстве).

– Часто постконтактная профилактика сочетает введение вакцины (для долгосрочного иммунитета) и готовых антител (иммуноглобулинов) для «пассивной» немедленной защиты.

Долгое время считалось, что вакцинировать уже больного человека бессмысленно, однако современные подходы **терапевтической вакцинации** пересматривают эту парадигму, особенно в онкологии и при хронических инфекциях [42]. Цель здесь – мобилизовать организм на борьбу с уже существующим патологическим процессом и устранение осложнений.

– Например, при гепатите B или C, ВИЧ-инфекции. Экспериментальные терапевтические вакцины направлены на стимуляцию Т-клеточного иммунитета, чтобы уничтожить клетки, в которых вирус персистирует.

– Лечебные противоопухолевые вакцины создаются на основе антигенов опухолевых клеток конкретного пациента. Они «обучают» иммунную систему распознавать и атаковать раковые клетки, которые обычно умеют «прятаться» от иммунного надзора [43, 44].

4 СТРАТЕГИЯ КУМУЛЯТИВНЫХ ДЕЦИМОДИФИКАЦИЙ

В классическом правиле Парето 20% усилий дают 80% результата, но данное соотношение нашло место в различных областях и как маркер эффективного распределения [45]. В контексте нуклеотидных последовательностей (ДНК/РНК) прямое применение этого правила затруднено из-за эффекта масштаба.

Как известно, в процессе его снижения начинают проявляться особенности и новые физические явления, которые не могут быть объяснены привычными законами физики, при этом точкой перехода является субатомный уровень. Именно поэтому при уменьшении физического масштаба до уровня отдельных нуклеотидов и пар оснований, «золотое сечение» эффективности стоит сместить к соотношению 10/90.

Таким образом, **децимодификацией** является изменение до одной десятой части целевого участка или локуса, что также позволяет применять итеративный (шаг за шагом) подход и производить более частые валидации (проверки).

Базовый цикл можно представить следующим образом:

1. Имеется нуклеотидная последовательность длиной N.
2. Определяется участок, ответственный за функцию (например, участок длиной L, который нужно изменить).
3. Вместо того чтобы сразу заменить весь участок L, выбирается фрагмент длиной $0.1 \times L$ (деци-часть).
4. Внесение точечных замен именно в эту «деци-часть».
5. Проверка, не разрушена ли общая структура. Если изменения прошли успешно, эта «деци-часть» фиксируется, а целевой участок L переопределяется (уменьшается на внесенные изменения).
6. Желательная сумма таких изменений должна быть не более 10% на каждом уровне, т.е. $0.1N$, $0.1L$ и т.д.

Кумулятивный эффект от таких изменений возникает за счет нескольких факторов, которые проявляются именно благодаря малости каждого отдельного изменения.

Во-первых, децимодификации позволяют обойти защитные механизмы клетки: замена 10% последовательности – это шаг с минимальным энергетическим порогом, тогда как крупные правки вызывают термодинамическую нестабильность. Серия таких шагов дает ту же дистанцию, что и радикальное редактирование, но без отторжения системой.

Во-вторых, малые правки работают синергично: каждая замена подстраивает контекст (стабильность, специфичность) под предыдущие. В итоге сумма изменений дает эффект выше номинального за счет оптимизации промежуточных состояний.

В-третьих, при использовании CRISPR поэтапное внесение правок сохраняет гомологию, необходимую для гибридизации: на ранних итерациях связывание идет по старому участку, на поздних – по уже сформированному новому консервативному участку.

Т.е **кумулятивный эффект** это превышение итогового результата как над арифметической суммой отдельных малых изменений, так и крупной модификацией.

Таким образом можно дать полное определение:

Стратегия кумулятивных децимодификаций – форма редактирования нуклеотидной последовательности, при которой каждое множественное вмешательство ограничено изменением десятой части от целевой области, а их кумулятивный эффект достигается за счет преодоления барьеров и итеративности.

Один из вариантов, как математически можно описать реализацию:

Представим, что у нас есть ген длиной 1000 п.н. (пар оснований). Наша цель – полностью изменить его функцию на новую. «Идеальная» целевая последовательность отличается от исходной на 800 п.н. (80% генома). Это большая дистанция.

Шаг 1: остаток неизмененной целевой области $R_0 = 800$ п.н. (это те позиции, которые в итоге должны стать другими). Мы берем 10% от 800 = 80 п.н.. Вносим 80 замен. Проверяем. Белок работает? Да (за счет того, что изменения малы, структура не рухнула).

Шаг 2: мы уже изменили 80 п.н. Значит, нам осталось изменить еще $800 - 80 = 720$ п.н. (цель не достигнута). Берем 10% от текущего остатка (720) = 72 п.н..

Шаг 3: остаток $720 - 72 = 648$. Берем 10% = 64,8 п.н**.** (округляем до 65). И т.д.

Если бы мы меняли каждый раз фиксированные 80 п.н. (10% от исходной длины), мы бы не приблизились к цели аккуратно, потому что остаток быстро стал бы меньше 80, и мы бы ломали лишнее. Формула $0.1 \times R_{n-1}$ гарантирует, что размер вмешательства пропорционален сложности оставшейся задачи. Чем ближе мы к идеалу, тем мельче шаги, тем выше точность.

Если мы просто сложим все эти децимодификации количественно (сколько нуклеотидов мы физически тронули рукой редактора), то получится ряд:

$$S = 80 + 72 + 64.8 + 58.3 + 52.5 + \ldots$$

Это геометрическая прогрессия. Если мы будем итерировать, пока остаток не станет близок к нулю, сумма всех изменений будет стремиться к 800 (к полному объему целевых замен). То есть, номинально мы сделали работу ровно на 100% от запланированной. Никакого чуда здесь нет: чтобы заменить 800 позиций, нужно в сумме сделать 800 замен.

Откуда тогда берется кумулятивный эффект (больше, чем крупное изменение), если замен мы сделали столько же?

Эффект возникает не только от количества замен, а и от качества их распределения и контекстного коэффициента C_n

Представим, что эффективность работы фермента (или стабильность РНК) измеряется в условных единицах от 0 до 100.

Исходная последовательность дает эффективность 20.

Целевая последовательность дает эффективность 100.

Случай А. Крупное изменение, мы берем и заменяем сразу 800 нуклеотидов. Эффект $= 100$ (цель), но Вероятность $= 0$. Белок не сворачивается, РНК не транслируется. Мы получили 0. Последовательность мертва. С точки зрения функции, мы не получили ничего, несмотря на 800 замен.

Случай Б. Децимодификации. У нас есть функция $E(S)$, которая описывает, как последовательность дает эффективность. Эта функция нелинейна.

Каждая следующая децимодификация дает прирост эффективности не на фиксированную величину, а на величину, пропорциональную текущему состоянию системы.

$$\Delta E_n = (0.1 \times R_{n-1}) \times k \times (1 + C_n) \quad (4.1)$$

где:

$0.1 \times R_{n-1}$ – количество измененных нуклеотидов.

k – базовый коэффициент влияния одной замены (например, 0.1% эффективности на замену).

C_n – коэффициент контекстной адаптации (синергии).

Как работает C_n:

На 1-м шаге $C_1 = 0$ (контекст еще не подготовлен). Мы изменили 80 п.н. Эффективность выросла с 20 до 28 (прибавили 8 единиц).

На 2-м шаге мы изменили 72 п.н. Но благодаря тому, что 80 п.н. уже изменены, структура молекулы слегка деформировалась, стала более «податливой» к следующим заменам. Вторая партия замен дает прирост не 7.2 единицы (как по линейному расчету), а 9 единиц. C_2 здесь = 0.25.

На 5-м шаге большая часть «жесткого каркаса» уже изменена, оставшиеся замены приходятся на гибкие участки. Одна замена на 5-м шаге может давать в 2 раза больше прироста эффективности, чем одна замена на 1-м шаге.

5 БИОИНФОРМАЦИОННОЕ МОДЕЛИРОВАНИЕ

После того как были рассмотрены основные аспекты и сформулирована исходная стратегия, перед нами встает задача ее объективной валидации. Наиболее релевантным способом такой проверки в контексте исследования представляется проведение вычислительного эксперимента в рамках биоинформационного подхода над конкретным объектом-инструментом.

5.1 Вирус Зика (ZIKV)

Вирус Зика – это арбовирусная инфекция, передаваемая преимущественно через укусы комаров рода Aedes, в первую очередь Aedes aegypti, которые также являются переносчиками лихорадки денге и чикунгуньи. Впервые вирус был выделен в 1947 году в лесу Зика в Уганде у макак-резусов, откуда и получил свое название, однако долгое время болезнь оставалась малоизвестной и вызывала лишь спорадические случаи в экваториальной Африке и Азии [45]. Ситуация кардинально изменилась в 2007 году, когда крупная вспышка произошла на острове Яп в Микронезии, а затем, начиная с 2013 года, вирус стремительно распространился по островам Тихого океана [46]. Подлинно глобальный резонанс заболевание получило в 2015 году, когда Бразилия столкнулась с беспрецедентной эпидемией, сопровождавшейся пугающим ростом числа новорожденных с микроцефалией, что заставило Всемирную организацию здравоохранения объявить чрезвычайную ситуацию международного масштаба [47].

Клиническое течение лихорадки Зика у большинства инфицированных людей либо протекает бессимптомно, либо ограничивается легкими гриппоподобными симптомами: повышением температуры, макулопапулезной сыпью (мелкими красными пятнами), конъюнктивитом, болями в суставах и мышцах, а также выраженной слабостью (см. рисунок 5.1) [48]. Эти проявления обычно длятся от двух до семи дней, и тяжелые формы, требующие госпитализации, встречаются крайне редко. Однако главная опасность вируса заключается не в его остром периоде, а в отсроченных последствиях. Именно в Бразилии была установлена прямая этиологическая связь между вирусом Зика и врожденными аномалиями плода: при инфицировании беременной женщины вирус обладает выраженным нейротропным действием, преодолевает плацентарный барьер и атакует нервные клетки развивающегося эмбриона, приводя к микроцефалии, внутримозговым кальцификатам и другим тяжелым порокам развития, объединенным термином «врожденный синдром, ассоциированный с вирусом Зика» [49].

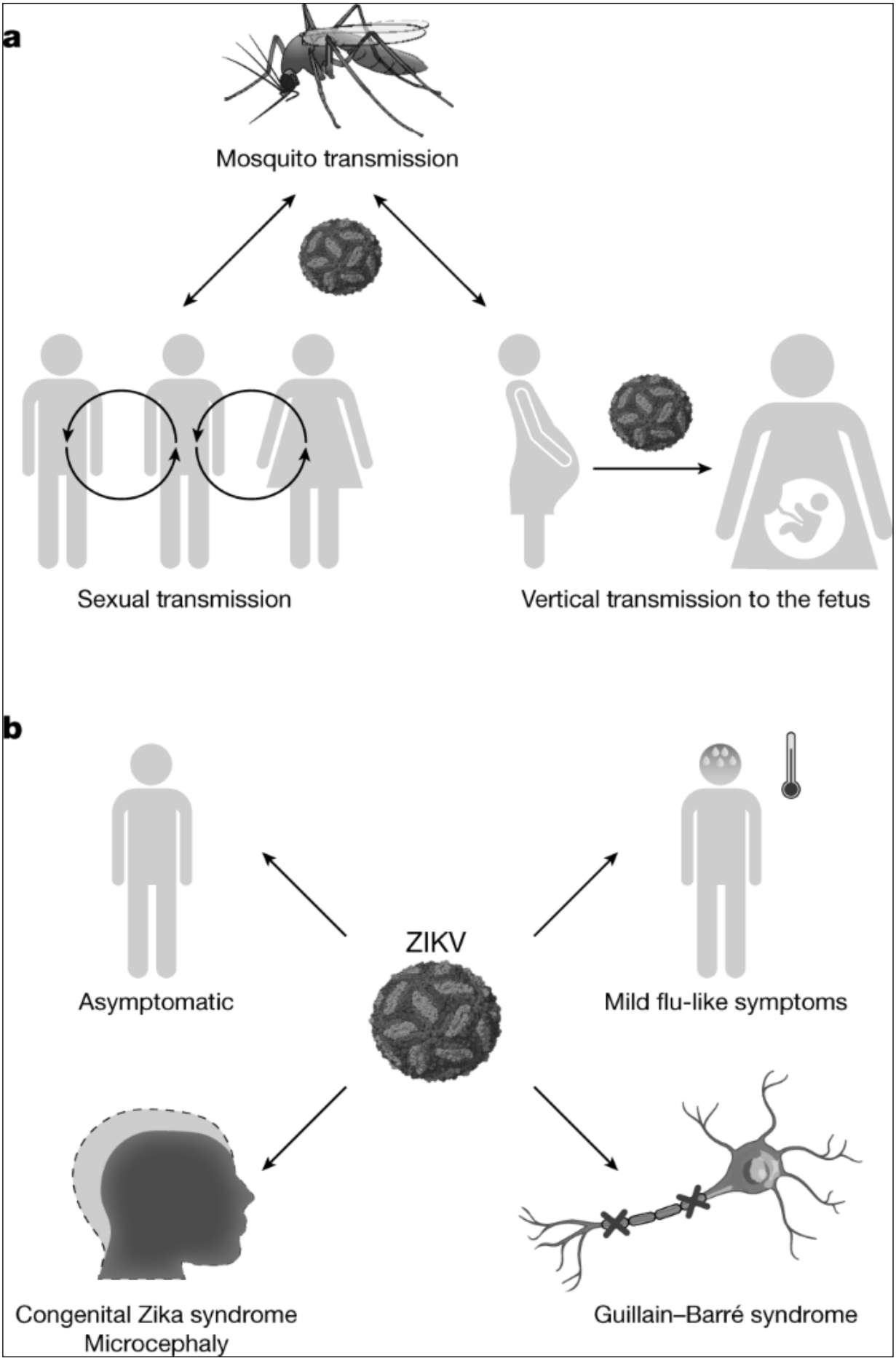

Рисунок 5.1 – Вирус Зика

Помимо трансмиссивного пути передачи (через комаров), вирус Зика выделяется среди прочих арбовирусов способностью передаваться половым путем, что является уникальным явлением для данной группы инфекций: вирус может сохраняться в семенной жидкости значительно дольше, чем в крови [50]. Также описаны случаи передачи при гемотрансфузиях, от матери к ребенку во время родов и, предположительно, через слюну, хотя последний путь не является эпидемиологически значимым. Диагностика заболевания осложняется перекрестными реакциями с другими флавивирусами, в частности с вирусом денге, поэтому для подтверждения диагноза применяется полимеразная цепная реакция в первые дни болезни, а позднее – серологические тесты, требующие подтверждения реакцией нейтрализации. Специфического этиотропного лечения, равно как и коммерчески доступной вакцины, на данный момент не существует, а терапия сводится к купированию симптомов, обильному питью и отдыху [51].

В связи с отсутствием вакцины основой борьбы с вирусом Зика является профилактика, которая включает два ключевых направления. Первое – это контроль численности комаров-переносчиков: уничтожение мест их размножения (стоячие водоемы, емкости с водой), использование инсектицидов и внедрение биологических методов борьбы, таких как заражение комаров бактерией Wolbachia, препятствующей размножению вируса. Второе направление – индивидуальная защита: применение репеллентов, ношение закрытой одежды, установка противомоскитных сеток, особенно в дневное время, когда комары Aedes наиболее активны. Беременным женщинам настоятельно рекомендуется избегать поездок в регионы с активной передачей вируса, а их половым партнерам, возвращающимся из эндемичных зон, – использовать барьерные методы контрацепции на протяжении нескольких месяцев. Хотя пик пандемии 2015–2016 годов прошел, вирус Зика продолжает циркулировать в тропических и субтропических регионах, оставаясь скрытой угрозой, способной в любой момент дать новую вспышку в неиммунном населении [51].

С точки зрения нуклеотидной последовательности вирус Зика представляет собой одноцепочечную РНК положительной полярности, длина которой составляет примерно 10 794 нуклеотида (точное значение может незначительно варьироваться в зависимости от штамма, например, у азиатского генотипа, вызвавшего пандемию в Америке, длина составляет 10 807 нуклеотидов) [52]. Эта молекула РНК выполняет сразу две функции: она является не только генетическим материалом, но и матрицей для синтеза белка, то есть фактически работает как иРНК. На 5'-конце генома находится кап, необходимый для инициации трансляции в рибосоме клетки хозяина, тогда как 3'-конец лишен полиаденилового хвоста, однако формирует характерные шпилечные структуры, стабилизирующие молекулу.

Весь геном вируса Зика организован в одну протяженную открытую рамку считывания, фланкированную двумя нетранслируемыми областями: 5'-NTR и 3'-NTR длиной около 100 и 400 нуклеотидов соответственно. Эти некодирующие участки содержат высококонсервативные вторичные структуры РНК, которые служат сигналами для репликации и упаковки генома. Внутри единственной рамки считывания закодирован полипротеин-предшественник, состоящий из 3419 аминокислотных остатков. Котрансляционно и посттрансляционно этот гигантский полипептид разрезается вирусной и клеточными протеазами на десять зрелых белков, порядок расположения которых строго фиксирован вдоль цепи РНК.

Структурные белки кодируются в 5'-трети генома в следующей последовательности: капсидный белок C (ответственный за упаковку РНК), белок prM (предшественник мембранного белка, необходимый для правильного созревания вириона) и белок E (белок оболочки). Именно в нуклеотидной последовательности гена E закодирована способность вируса связываться с рецепторами клеток и проникать внутрь, а мутации в этом участке часто определяют нейротропность вируса [53, 54]. Оставшиеся две трети генома кодируют семь неструктурных белков: NS1, NS2A, NS2B, NS3, NS4A, NS4B и NS5. Среди них ключевыми ферментами являются NS5, содержащий РНК-зависимую РНК-полимеразу, которая обеспечивает копирование вирусной РНК, и NS3, который в комплексе со своим кофактором NS2B функционирует как сериновая протеаза, разрезающая полипротеин. Нуклеотидная последовательность участка NS1 интересна тем, что кодирует белок, способный ингибировать систему комплемента человека, а также служит одним из главных антигенов при серологической диагностике.

Филогенетический анализ нуклеотидных последовательностей позволил разделить вирус Зика на два основных генотипа: африканский и азиатский. Азиатский генотип, вызвавший вспышку в Бразилии, имеет характерные однонуклеотидные замены по сравнению с африканскими предками. Особое внимание исследователей привлекла миссенс-мутация в гене NS1, приводящая к замене серина на аспарагин в позиции 139 (S139N). Ретроспективный анализ показал, что эта нуклеотидная замена появилась в азиатской линии около 2013 года и статистически достоверно ассоциирована с повышенным риском развития микроцефалии. Считается, что данная замена усилила способность вируса инфицировать клетки нейронального происхождения.

Кроме того, при анализе нуклеотидного состава вируса Зика обращает на себя внимание низкое содержание GC-пар (гуанин-цитозин) по всему геному, что характерно для флавивирусов, адаптированных к комарам. Это проявляется в частом использовании кодонов с аденином и урацилом на третьей позиции. Интересно, что, несмотря на свою

РНК-природу и высокую частоту мутаций, характерную для РНК-вирусов в целом, вирус Зика демонстрирует относительно низкую скорость нуклеотидных замен по сравнению с вирусом денге или Западного Нила, что объясняется жесткими ограничениями, накладываемыми необходимостью сохранения вторичных структур РНК в 3'-нетранслируемой области. Эти структуры жизненно важны для инициации синтеза минус-цепи РНК, и даже одиночные нуклеотидные замены, нарушающие спаривание оснований в шпильках, часто оказываются летальными [54].

Говоря о нуклеотидной последовательности, нельзя не упомянуть проблему рекомбинации. Хотя для флавивирусов рекомбинация считается редким событием, анализ полных геномов вируса Зика выявил следы гомоплазий, то есть одинаковых нуклеотидных замен, возникших независимо в разных эволюционных ветвях. Это создает трудности для филогенетической реконструкции и указывает на то, что вирус способен приходить к одним и тем же адаптивным решениям разными эволюционными путями.

5.2 Анализ последовательности и выбор области редактирования

На первом этапе для анализа нуклеотидной последовательности используем библиотеку NCBI, где отберем имеющиеся варианты и штаммы вируса Зика (в сумме девять, см. приложение А).

Рисунок 5.2 – Пример ввода и результатов отбора последовательностей

Вторым этапом произведем их множественное выравнивание с помощью сервиса Clustal Omega. При этом после первого раза на древе видно, что варианты 5 и 9 отличны от остальных. Для лучшей однородности исключим их и повторим выравнивание.

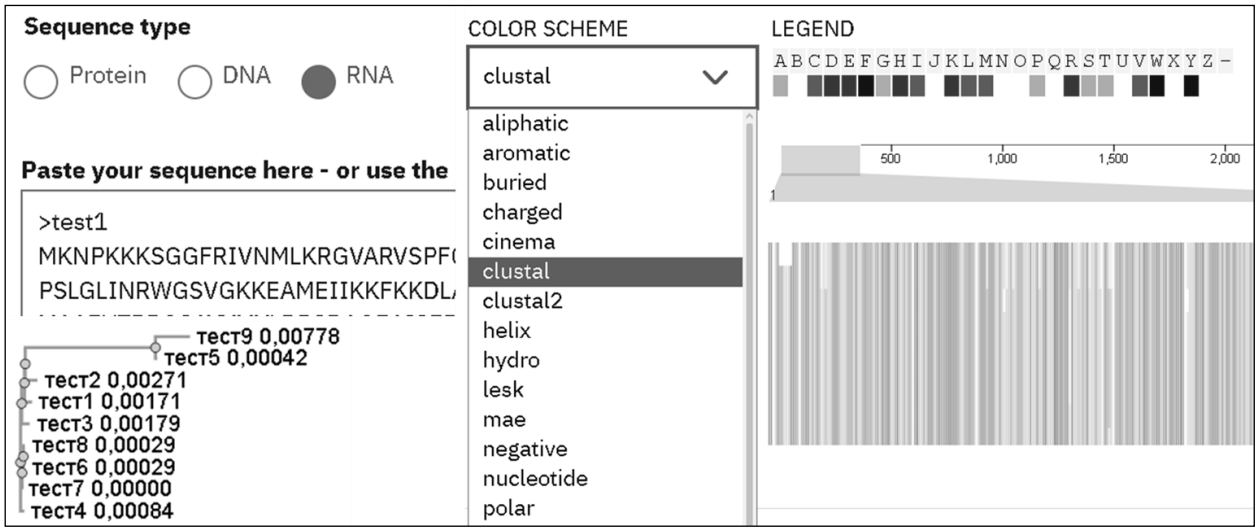

Рисунок 5.3 – Пример ввода и результаты первого выравнивания

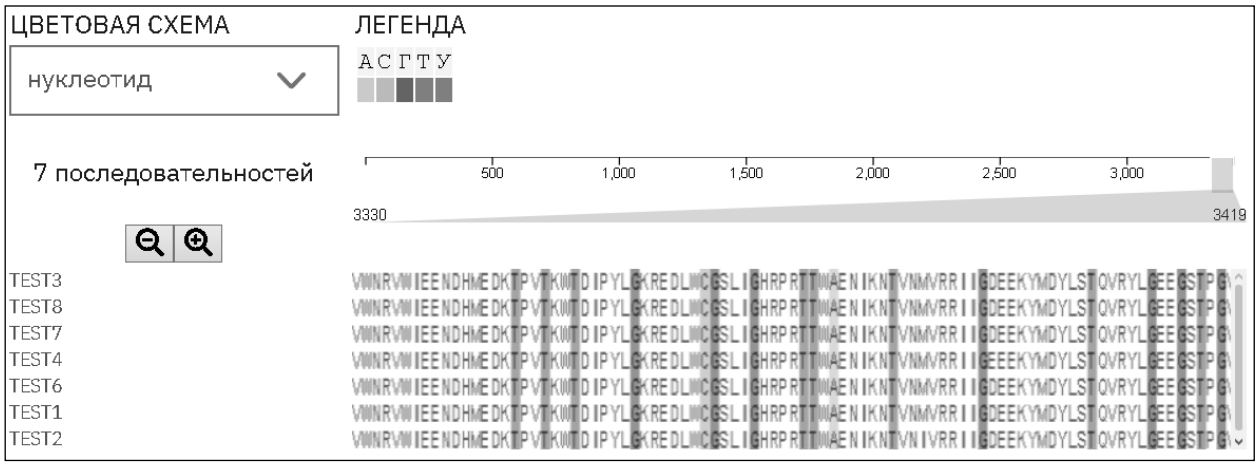

Рисунок 5.4 – Результаты второго выравнивания

По результатам второго выравнивания получаем визуализацию и возможность выбирать консервативные области, которые будут релевантны для всех вариантов вируса.

Следующим этапом является выбор участка, в котором будут производиться модификации. Исходя из данных раздела про вакцины и описания вируса Зика, при создании вакцины на основе вирусного вектора мы должны ориентироваться на три параметра: остановка репликации, остановка противодействия иммунитету, модуляция восстановления. Для каждого можно подобрать участки, редактирование которого обеспечит решение задачи:

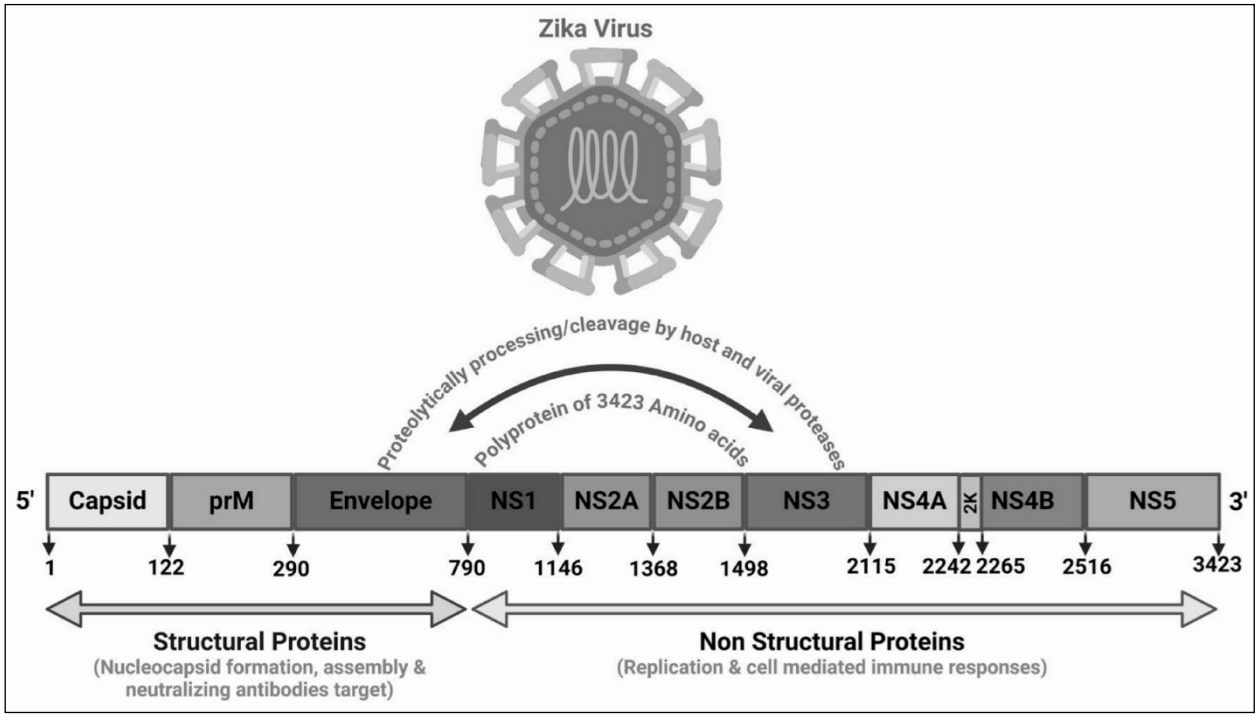

Рисунок 5.5 – Участки генома вируса Зика

– Остановка репликации вируса

Ген NS5 (участок РНК-полимеразы)

Примерно 9000–10 000 нуклеотид.

Это РНК-зависимая РНК-полимераза. Если этот белок не соберется или будет синтезирован с ошибкой, вирус физически не сможет размножаться.

Ген NS3 (хеликазный домен)

Примерно 5200–6200 нуклеотид.

NS3 расплетает двухцепочечную РНК. Без него полимераза (NS5) не сможет прочитать матрицу.

3'-Нетранслируемая область (3'-UTR)

Концевые 100 нуклеотидов (~10 694–10 794).

Удаление промоторных шпилек. Это цис-элемент (действует только на той РНК, где находится)

– Остановка противодействия иммунитету

Ген NS4B

Примерно 7200–7600 нуклеотид.

NS4B вируса Зика — мощный ингибитор интерферона. Он блокирует фосфорилирование STAT1/STAT2 и нарушает сигнальный путь JAK-STAT.

Инфицированная клетка начнет секретировать интерфероны и звать Т-киллеры.

Ген NS5 (N-концевой домен метилтрансферазы)

Примерно 7600–8200 нуклеотид.

NS5 флавивирусов связывается с белком STAT2 человека и вызывает его деградацию. Без этого домена вирусная РНК становится иммуногенной (клетка распознает её как чужеродную).

Ген NS2A

Примерно 3300–3800 нуклеотид.

Участвует в подавлении интерферонового ответа на ранних стадиях. Его инактивация снижает вирулентность

– Модуляция восстановления

Вставка в 3'-UTR

Сразу после стоп-кодона NS5, перед конечной шпилькой 3'-UTR

до ~300 нуклеотидов

отдельный промотор

Замена гена prM-E и т.д.

Но возникает препятствие. Если мы полностью остановим репликацию, то производство станет сложным и дорогим; если мы сохраним репликацию, но попытаемся ослабить противодействие иммунитету, то мы не решим проблему его опасности, а сам вирус сможет откатиться (мутировать) в исходную или более опасную форму; добавлять в любом из этих случаев модулирующие участки нерационально, т.к. какой смысл пытаться восстановить, если вирус будет продолжать разрушать.

Мы приходит к тому, что необходимо воспользоваться стратегией кумулятивных децимодификаций более радикально и не просто изменять до 10% области последовательности, но полностью убрать другие 90. Таким образом, вместо использования самого вируса Зика в качестве вектора, будут использованы лишь его участки prM и E.

Далее, с использованием сервисов CHOPCHOP и Cas-Offinder находим целевые и нецелевые сайты, что позволит нам произвести замены, к примеру для ухода от антител (EDIII (домен III белка E; замены аминокислот (например, T335R, T337R, D389L)) или снижения нейротропизма (сайт S139N; обратная мутация N139S) и т.д.

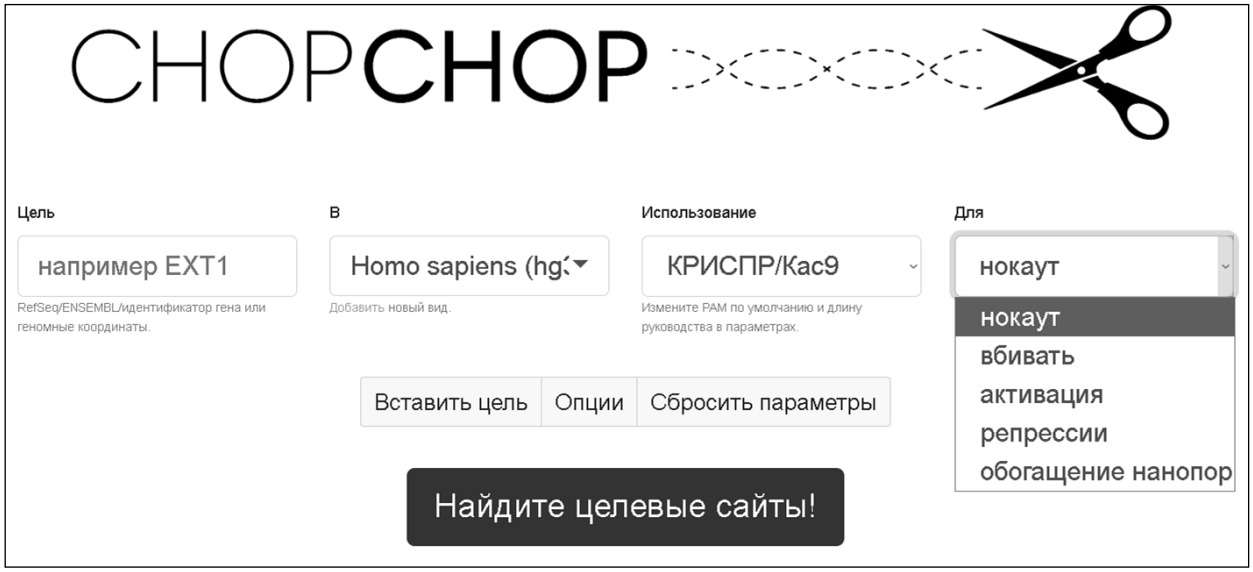

Рисунок 5.6 – Меню ввода сервиса CHOPCHOP

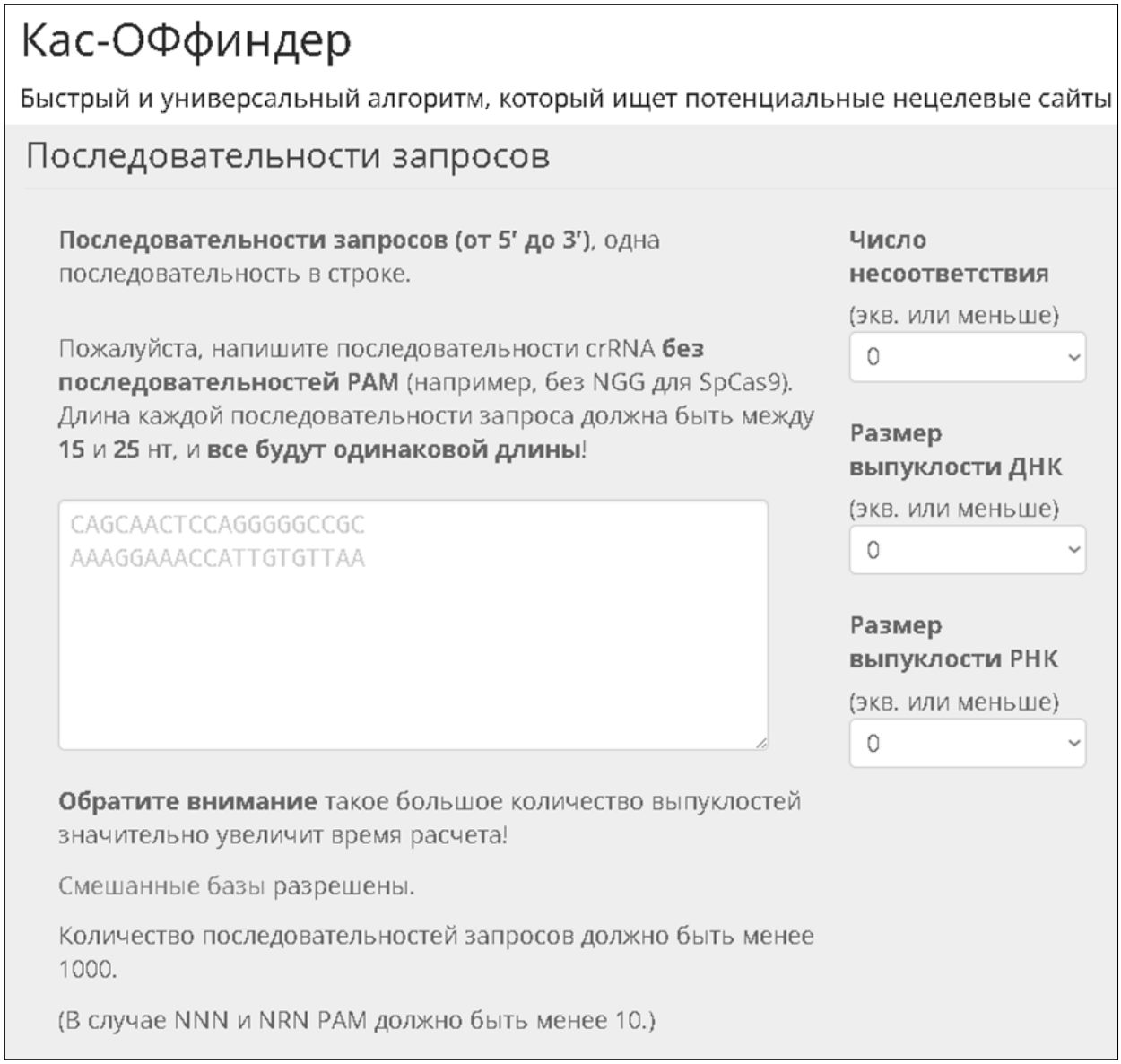
Кас-ОФФиндер
Быстрый и универсальный алгоритм, который ищет потенциальные нецелевые сайты
Последовательности запросов
Последовательности запросов (от 5' до 3'), одна последовательность в строке.
Пожалуйста, напишите последовательности crRNA без последовательностей PAM (например, без NGG для SpCas9). Длина каждой последовательности запроса должна быть между 15 и 25 нт, и все будут одинаковой длины!
CAGCAACTCCAGGGGGCCGC
AAAGGAAACCATTGTGTTAA
Обратите внимание такое большое количество выпуклостей значительно увеличит время расчета!
Смешанные базы разрешены.
Количество последовательностей запросов должно быть менее 1000.
(В случае NNN и NRN PAM должно быть менее 10.)
Число несоответствия
(экв. или меньше)
0
Размер выпуклости ДНК
(экв. или меньше)
0
Размер выпуклости РНК
(экв. или меньше)
0

Рисунок 5.7 – Меню ввода сервиса Cas-Offinder

5.3 Построение вирусного вектора

В итоге мы можем построить модель-схему конструкта на основе вируса Зика, указав целевые участки и их место расположения.

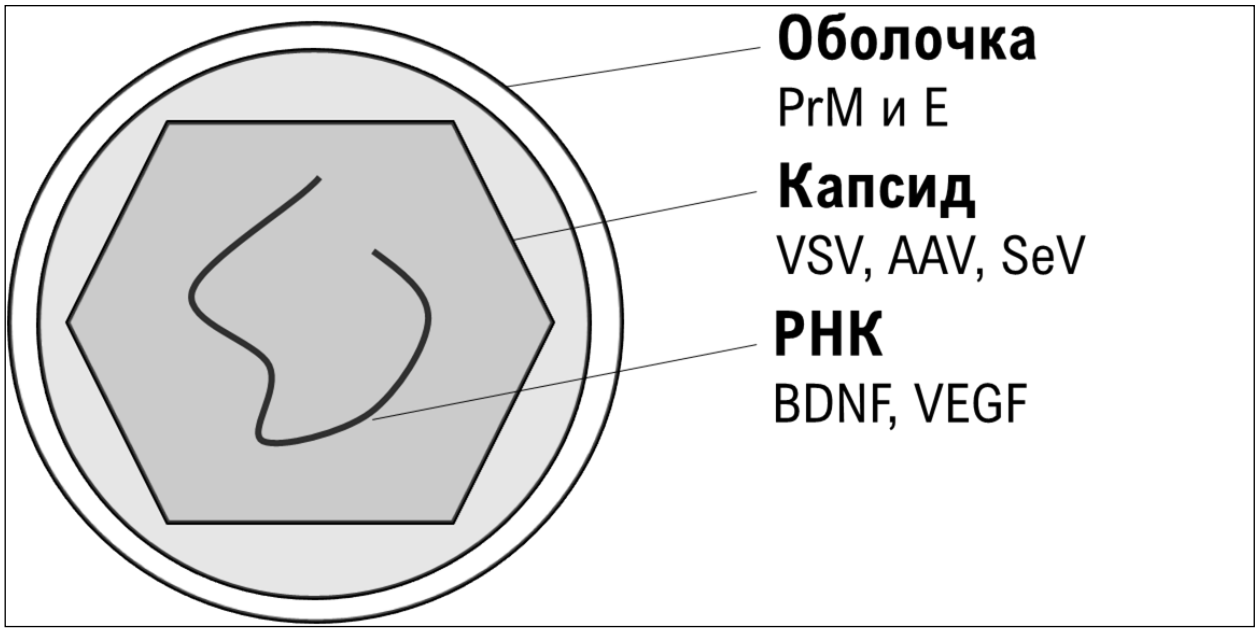

Рисунок 5.8 – Модель-схема вирусного вектора (ВСТАВИТЬ СХЕМУ)

В оболочке используем выбранные и измененные участки вируса Зика; в качестве носителя-капсида используем технологичные варианты, т.е. производим псевдотипирование; в РНК встраиваем последовательности с фокусом на нервную и капиллярные системы.

На этом этапе работа по моделированию является незаконченной, о чем далее.

6 ПРИМЕНЕНИЕ В РАМКАХ ПРОЕКТА

Проект – целенаправленное использование ресурсов и элементов внешней среды для достижения заранее обозначенного результата [55]. Опишем гипотетический цикл проекта по созданию вирусного вектора, а затем детализирем этапы, фактически реализованные в данной диссертации.

1. **Выбор целевой биомедицинской задачи.** Определение, для чего именно создается вектор: для доставки терапевтического гена, в качестве ослабленной живой вакцины или для редактирования генома клеток хозяина.
2. **Выбор исходного штамма и анализ его генома.** Выбор референтного или клинического изолята.
3. **Биоинформационное моделирование целевых модификаций.** Определение участков генома, ответственных за ключевые функции.
4. **Разработка конструкта вектора.** Выбор типа вносимых изменений (нуклеотидные замены, делеции, инсерции) и их пространственно-временной последовательности.
5. **Валидация полученной последовательности.**
6. **Лабораторный синтез и сборка.** Химико-ферментативный синтез модифицированных фрагментов РНК/ДНК и сборка полноразмерного генома вектора (например, с использованием методов клонирования).
7. **Трансфекция и получение вирусных частиц.** Введение сконструированного генома в клеточные линии для получения полноценных вирионов.
8. **Валидация in vitro.** Проверка способности вектора инфицировать целевые клетки, его репликативной активности и отсутствия цитопатического эффекта.
9. **Доклинические исследования in vivo.** Оценка безопасности, иммуногенности и терапевтической эффективности на животных моделях (мышки, крыски, приматы).
10. **Клинические исследования.** Поэтапная проверка на людях (от фазы I до фазы III).
11. **Регистрация и производство.** Масштабирование технологии для клинического применения.

В настоящей работе, в силу её теоретико-методологического характера, реализованы этапы 1–3 и частично этап 4 из вышеуказанного цикла.

Этапы лабораторного синтеза, сборки и валидации такого вектора составляют перспективу дальнейших исследований. Тем не менее, заложенная теоретическая и биоинформационная база является фундаментом для их проведения.

РЕЗУЛЬТАТЫ И ВЫВОДЫ

В ходе выполнения данной работы были получены следующие основные **результаты**, соответствующие поставленной цели и задачам исследования:

1. Создание теоретической модели вирусного вектора.

На основе проведенного анализа и с использованием сервисов для поиска мишеней редактирования (CHOPCHOP, Cas-Offinder) теоретически сконструирован гипотетический вирусный вектор. В отличие от использования полного генома вируса, стратегия КДМ применена для создания конструкции на основе фрагментов генов *prM* и *E*, что позволяет переориентировать цель только с создания ослабленной вакцины на разработку безопасной платформы доставки. Для данной модели были определены целевые сайты модификации, что подтверждает принципиальную возможность направленного изменения свойств биологического агента без потери его функциональности.

2. Обоснование стратегии кумулятивных децимодификаций (КДМ).

Впервые предложена и теоретически обоснована стратегия редактирования нуклеотидных последовательностей, основанная на принципе итеративного внесения малых изменений. Показано, что ограничение размера единовременной модификации до 10% от целевого участка (децимодификация) позволяет минимизировать дестабилизацию молекулярной структуры и избегать запуска летальных для клетки защитных механизмов (апоптоза, неконтролируемой репарации). Сформулирована математическая модель (4.1), описывающая нелинейный (кумулятивный) прирост функциональной эффективности за счет контекстной адаптации последовательности на каждом этапе.

3. Разработка биоинформационного подхода к валидации стратегии.

На примере вируса Зика (ZIKV) продемонстрирована применимость биоинформационного моделирования (*in silico*) для проверки эффективности стратегии КДМ. Проведен анализ нуклеотидных последовательностей девяти штаммов ZIKV с использованием методов множественного выравнивания (Clustal Omega), что позволило идентифицировать консервативные и вариабельные участки генома, ответственные за ключевые функции патогена (репликация, уход от иммунного ответа, нейротропизм).

На основе полученных результатов можно сформулировать следующие **выводы**:

1. Биоинформационное моделирование является высокоэффективным методом научного познания в области биотехнологии.

2. Технологии редактирования нуклеотидной последовательности представляют собой перспективную область исследований с широкими рамками для разработок.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ И ЛИТЕРАТУРЫ

1. Захаров А.М. (2025). Развитие природной метасистемы на основе биохомоценоза. Магистерская диссертация. https://doi.org/10.13140/RG.2.2.11353.38246
2. Альбертс Б., Брей Д., Хопкин К. и др. «Основы молекулярной биологии клетки» (5-е издание и новее). – М.: Лаборатория знаний, 2022.
3. Юрина Н. А., Радостина А. И. «Гистология, цитология и эмбриология» (учебник). – М.: Медицина, 2019.
4. Льюин Б., Кассмерис Л., Лингаппа В.Р., Плоппер Г. «Клетки» (3-е издание). – М.: Лаборатория знаний, 2022.
5. Жимулев И.Ф. «Общая и молекулярная генетика» (учебное пособие). – Новосибирск: Изд-во НГУ, 2017
6. Нельсон Д.Л., Кокс М.М. «Принципы биохимии Ленинджера» (7-е издание и новее). – М.: Лаборатория знаний, 2022.
7. Маррилли А., Дерёж П., Ритц Ж., Турнель А. «Геномика и биоинформатика» (пер. с франц.). – М.: Лаборатория знаний, 2020.
8. Strachan, T. Human Molecular Genetics / T. Strachan, A. Read. – 5th ed. – Boca Raton : CRC Press, 2018. – 770 p.
9. Ingram, V. M. A specific chemical difference between the globins of normal human and sickle-cell anaemia haemoglobin / V. M. Ingram // Nature. – 1956. – Vol. 178, No. 4537. – P. 792–794. – DOI: 10.1038/178792a0.
10. Riordan, J. R. Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA / J. R. Riordan [et al.] // Science. – 1989. – Vol. 245, No. 4922. – P. 1066–1073. – DOI: 10.1126/science.2475911.
11. Bobadilla, J. L. Cystic fibrosis: a worldwide analysis of CFTR mutations—correlation with incidence data and application to screening / J. L. Bobadilla [et al.] // Human Mutation. – 2002. – Vol. 19, No. 6. – P. 575–606.
12. Myerowitz, R. Tay-Sachs disease-causing mutations and neutral polymorphisms in the Hex A gene / R. Myerowitz // Human Mutation. – 1997. – Vol. 9, No. 3. – P. 195–208.
13. The Huntington's Disease Collaborative Research Group. A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes // Cell. – 1993. – Vol. 72, No. 6. – P. 971–983. – DOI: 10.1016/0092-8674(93)90585-e.
14. Verkerk, A. J. Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome / A. J. Verkerk [et al.] // Cell. – 1991. – Vol. 65, No. 5. – P. 905–914.

15. Blau, N. Phenylketonuria / N. Blau // Nature Reviews Disease Primers. – 2021. – Vol. 7, No. 1. – P. 36. – DOI: 10.1038/s41572-021-00267-0.
16. Veitia, R. A. Exploring the etiology of haploinsufficiency / R. A. Veitia, J. A. Birchler // BioEssays. – 2002. – Vol. 24, No. 2. – P. 175–184.
17. Finkbeiner, S. Huntington's Disease and the Neurodevelopmental Hypothesis / S. Finkbeiner // Neuron. – 2019. – Vol. 101, No. 5. – P. 780–782.
18. Vogelstein, B. The Genetic Basis of Human Cancer / B. Vogelstein, K. W. Kinzler. – 2nd ed. – New York : McGraw-Hill, 2002. – 887 p.
19. Alberts, B. Молекулярная биология клетки / B. Alberts [et al.]; пер. с англ. под ред. Г. П. Георгиева. – 6-е изд. – Москва : Лаборатория знаний, 2022. – (Глава 20: "Рак").
20. Brash, D. E. UV signature mutations / D. E. Brash // Photochemistry and Photobiology. – 2015. – Vol. 91, No. 1. – P. 15–26. – DOI: 10.1111/php.12377.
21. Pfeifer, G. P. Tobacco smoke carcinogens, DNA damage and p53 mutations in smoking-associated cancers / G. P. Pfeifer [et al.] // Oncogene. – 2002. – Vol. 21, No. 48. – P. 7435–7451. – DOI: 10.1038/sj.onc.1205803.
22. Alexandrov, L. B. Mutational signatures associated with tobacco smoking in human cancer / L. B. Alexandrov [et al.] // Science. – 2016. – Vol. 354, No. 6312. – P. 618–622.
23. Davies, H. Mutations of the BRAF gene in human cancer / H. Davies [et al.] // Nature. – 2002. – Vol. 417, No. 6892. – P. 949–954. – DOI: 10.1038/nature00766.
24. Olivier, M. TP53 mutations in human cancers: origins, consequences, and clinical use / M. Olivier, M. Hollstein, P. Hainaut // Cold Spring Harbor Perspectives in Biology. – 2010. – Vol. 2, No. 1. – P. a001008.
25. Ionov, Y. Ubiquitous somatic mutations in simple repeated sequences reveal a new mechanism for colonic carcinogenesis / Y. Ionov [et al.] // Nature. – 1993. – Vol. 363, No. 6429. – P. 558–561.
26. Martincorena, I. Somatic mutation in cancer and normal cells / I. Martincorena, P. J. Campbell // Science. – 2015. – Vol. 349, No. 6255. – P. 1483–1489.
27. Naldini, L. Gene therapy returns to centre stage / L. Naldini // Nature. – 2015. – Vol. 526, No. 7573. – P. 351–360. – DOI: 10.1038/nature15818.
28. Doudna, J. A. The new frontier of genome engineering with CRISPR-Cas9 / J. A. Doudna, E. Charpentier // Science. – 2014. – Vol. 346, No. 6213. – P. 1258098.
29. National Academies of Sciences, Engineering, and Medicine. Human Genome Editing: Science, Ethics, and Governance / National Academies of Sciences, Engineering, and Medicine. – Washington, DC : The National Academies Press, 2017. – 328 p. – DOI: 10.17226/24623.

30. Almeida-Porada, G. In utero gene therapy: progress and challenges / G. Almeida-Porada, C. D. Porada // Molecular Therapy. – 2021. – Vol. 29, No. 10. – P. 2895–2908. – DOI: 10.1016/j.ymthe.2021.05.019.
31. Cartier, N. Hematopoietic stem cell gene therapy for lysosomal storage disorders: expectations and limitations / N. Cartier [et al.] // Hematology/Oncology Clinics of North America. – 2017. – Vol. 31, No. 5. – P. 861–875.
32. Mendell, J. R. Single-dose gene-replacement therapy for spinal muscular atrophy / J. R. Mendell [et al.] // The New England Journal of Medicine. – 2017. – Vol. 377, No. 18. – P. 1713–1722. – DOI: 10.1056/NEJMoa1706198.
33. June, C. H. CAR T cell immunotherapy for human cancer / C. H. June, M. Sadelain // Science. – 2018. – Vol. 359, No. 6382. – P. 1361–1365. – DOI: 10.1126/science.aar6711.
34. Tebas, P. Gene editing of CCR5 in autologous CD4 T cells of persons infected with HIV / P. Tebas [et al.] // The New England Journal of Medicine. – 2014. – Vol. 370, No. 10. – P. 901–910. – DOI: 10.1056/NEJMoa1300662.
35. Dunbar, C. E. Gene therapy comes of age / C. E. Dunbar [et al.] // Science. – 2018. – Vol. 359, No. 6372. – P. eaan4672. – DOI: 10.1126/science.aan4672.
36. Wang, D. Adeno-associated virus vector as a platform for gene therapy delivery / D. Wang, P. W. L. Tai, G. Gao // Nature Reviews Drug Discovery. – 2019. – Vol. 18, No. 5. – P. 358–378. – DOI: 10.1038/s41573-019-0012-9.
37. David, A. L. In utero gene therapy: current challenges and clinical perspectives / A. L. David, F. Peebles // Molecular Diagnosis & Therapy. – 2021. – Vol. 25, No. 4. – P. 425.
38. Pardridge, W. M. The blood-brain barrier: bottleneck in brain drug development / W. M. Pardridge // NeuroRx. – 2005. – Vol. 2, No. 1. – P. 3–14.
39. Pardi, N. mRNA vaccines — a new era in vaccinology / N. Pardi [et al.] // Nature Reviews Drug Discovery. – 2018. – Vol. 17, No. 4. – P. 261–279. – DOI: 10.1038/nrd.2017.243.
40. Plotkin, S. A. Correlates of protection induced by vaccination / S. A. Plotkin // Clinical and Vaccine Immunology. – 2010. – Vol. 17, No. 7. – P. 1055–1065. – DOI: 10.1128/CVI.00131-10.
41. WHO. Rabies vaccines: WHO position paper – April 2018 / World Health Organization // Weekly Epidemiological Record. – 2018. – Vol. 93, No. 16. – P. 201–220.
42. Finn, O. J. The dawn of vaccines for cancer / O. J. Finn // The Journal of Immunology. – 2018. – Vol. 200, No. 5. – P. 1561–1562.
43. Sahin, U. Personalized RNA mutanome vaccines mobilize poly-specific therapeutic immunity against cancer / U. Sahin [et al.] // Nature. – 2017. – Vol. 547, No. 7662. – P. 222–226. – DOI: 10.1038/nature23003.

44. Rosenberg, S. A. Cancer immunotherapy: a review / S. A. Rosenberg // The Journal of the American Medical Association. – 2014. – Vol. 312, No. 5. – P. 533–534.
45. Правило Парето / [Интернет источник] // URL: https://ru.wikipedia.org/wiki/Закон_Парето
46. Dick, G. W. Zika virus. I. Isolations and serological specificity / G. W. Dick, S. F. Kitchen, A. J. Haddow // Transactions of the Royal Society of Tropical Medicine and Hygiene. – 1952. – Vol. 46, No. 5. – P. 509–520. – DOI: 10.1016/0035-9203(52)90042-4.
47. Duffy, M. R. Zika virus outbreak on Yap Island, Federated States of Micronesia / M. R. Duffy [et al.] // The New England Journal of Medicine. – 2009. – Vol. 360, No. 24. – P. 2536–2543. – DOI: 10.1056/NEJMoa0805715.
48. Faria, N. R. Zika virus in the Americas: Early epidemiological and genetic findings / N. R. Faria [et al.] // Science. – 2016. – Vol. 352, No. 6283. – P. 345–349. – DOI: 10.1126/science.aaf5036.
49. Petersen, L. R. Zika virus / L. R. Petersen, D. J. Jamieson, A. M. Powers // The New England Journal of Medicine. – 2016. – Vol. 374, No. 16. – P. 1552–1563. – DOI: 10.1056/NEJMra1602113.
50. Mlakar, J. Zika virus associated with microcephaly / J. Mlakar [et al.] // The New England Journal of Medicine. – 2016. – Vol. 374, No. 10. – P. 951–958. – DOI: 10.1056/NEJMoa1600651.
51. Musso, D. Zika virus: from diagnosis to epidemic / D. Musso, D. J. Gubler // Clinical Microbiology Reviews. – 2016. – Vol. 29, No. 3. – P. 487–524.
52. World Health Organization. Zika virus: Fact sheet [Electronic resource] / World Health Organization. – 2022. URL: https://www.who.int/news-room/fact-sheets/detail/zika-virus
53. Faye, O. Molecular evolution of Zika virus during its emergence in the 20th century / O. Faye [et al.] // PLoS Neglected Tropical Diseases. – 2014. – Vol. 8, No. 1. – P. e2636. – DOI: 10.1371/journal.pntd.0002636.
54. Yuan, L. A single mutation in the prM protein of Zika virus contributes to fetal microcephaly / L. Yuan [et al.] // Science. – 2017. – Vol. 358, No. 6365. – P. 933–936. – DOI: 10.1126/science.aam7120.
55. Роль патентных исследований при реализации крупных инвестиционных и инженерных и проектов / А. А. Рягина, А. М. Захаров, Ю. И. Рягин. — Текст : электронный // Российские регионы в фокусе перемен : сборник докладов XIX Международной конференции (Екатеринбург, 14–16 ноября 2024 г.). — Екатеринбург : Издательство Издательский Дом «Ажур», 2025. — С. 402-405.

Приложение А

1. Zika virus ZIKV/Hu/Thai/KngSG/17-D501 genomic RNA, nearly complete genome

GenBank: LC369584.1

MKNPKKKSGGFRIVNMLKRGVARVSPFGGLKRLPAGLLLGHGPIRMVLAILAFLRFTAIKPSLGLINRWGSVGKKEAMEIIKKFKKDLA
AMLRIINARKEKKRRGADTSVGIVGLLLTTAMAAEVTRRGSAYYMYLDRSDAGEAISFPTTLGMNKCYIQIMDLGHMCDATMSYECP
MLDEGVEPDDVDCWCNTTSTWVVYGTCHHKKGEARRSRRAVTLPSHSTRKLQTRSQTWLESREYTKHLIRVENWIFRNPGFALAAAA
IAWLLGSSTSQKVIYLVMILLIAPAYSIRCIGVSNRDFVEGMSGGTWVDVVLEHGGCVTVMAQDKPTVDIELVTTTVSNMAEVRSYCYE
ASISDMASDSRCPTQGEAYLDKQSDTQYVCKRTLVDRGWGNGCGLFGKGSLVTCAKFACSKKMTGKSIQPENLEYRIMLSVHGSQHS
GMIVNDTGHETDENRAKVEITPNSPRAEATLGGFGSLGLDCEPRTGLDFSDLYYLTMNNKHWLVHKEWFHDIPLPWHAGADTGTPHW
NNKEALVEFKDAHAKRQTVVVLGSQEGAVHTALAGALEAEMDGAKGRLSSGHLKCRLKMDKLRLKGVSYSLCTAAFTFTKIPAETLH
GTVTVEVQYAGTDGPCKVPAQMAVDMQTLTPVGRLITANPVITESTENSKMMLELDPPFGDSYIVIGVGEKKITHHWHRSGSTIGKAFE
ATVRGAKRMAVLGDTAWDFGSVGGALNSLGKGIHQIFGAAFKSLFGGMSWFSQILIGTLLMWLGLNTKNGSISLMCLALGGVLIFLST
AVSADVGCSVDFSKKETRCGTGVFVYNDVEAWRDRYKYHPDSPRRLAAAVKQAWEDGICGISSVSRMENIMWRSVEGELNAILEENG
VQLTVVVGSVKNPMWRGPQRLPVPVNELPHGWKAWGKSYFVRAAKTNNSFVVDGDTLKECPLKHRAWNSFLVEDHGFGVFHTSVW
LKVREDYSLECDPAVIGTAVKGKEAVHSDLGYWIESEKNDTWRLKRAHLIEMKTCEWPKSHTLWTDGIEESDLIIPKSLAGPLSHHNTR
EGYRTQMKGPWHSEELEIRFEECPGTKVHVEETCGTGPSLRSTTASGRVIEEWCCRECTMPPLSFRAKDGCWYGMEIRPRKEPESNLVR
SMVTAGSTDHMDHFSLGVLVILLMVQEGLKKRMTTKIIISTSMAVLVAMILGGFSMSDLAKLTLIMGATFAEMNTGGDVAHLALIAAF
KVRPALLVSFIFRANWTPRESMLLALASCLLQTAISALEGDLMVLINGFALAWLAIRAMVVPRTDNITLAILAALTPLARGTLLVAWRA
GLATCGGFMLLSLKGKGSVKKNLPFVMALGLTAVRLVDPINVVGLLLLTRSGKRSWPPSEVLTAVGLICALAGGFAKADIEMAGPMA
AVGLLTVSYVVSGKSVDMYIERAGDITWEKDAEVTGNSPRLDVALDESGDFSLVEDDGPPMREIILKVVLMTICGMNPIAIPFAAGAWY
VYVKTGKRSGALWDVPAPKEVKKGETTDGVYRVMTRRLLGSTQVGVGVMQEGVFHTMWHVTKGSALRSGEGRLDPYWGDVKQDL
VSYCGPWKLDAAWDGHSEVQLLAVPPGERARNIQTLPGIFKTKDGDIGAVALDYPAGTSGSPILDKCGRVIGLYGNGVVIKNGSYVSAI
TQGRREEETPVECFEPSMLKKKQLTVLDLHPGAGKTRRVLPEIVREAIKTRLRTVILAPTRVVAAEMEEALRGLPVRYMTTAVNVTHSG
TEIVDLMCHATFTSRLLQPIRVPNYNLYIMDEAHFTDPSSIAARGYISTRVEMGEAAAIFMTATPPGTRDAFPDSNSPIMDTEVEPERAWS
SGFDWVTDHSGKTVWFVPSVRNGNEIAACLTKAGKRVIQLSRKTFETEFQKTKHQEWDFVVTTDISEMGANFKADRVIDSRRCLKPVI
LDGERVILAGPMPVTHASAAQRRGRIGRNPNKPGDEYLYGGGCAETDEDHAHWLEARMLLDNIHLQDGLIASLYRPEADKVAAIEGEF
KLRTEQRKTFVELMKRGDLPVWLAYQVASAGITYTDRRWCFDGTTNNTIMEDSVPAEVWTRYGEKRVLKPRWMDARVCSDHAALK
SFKEFAAGKRGAAFGVMEALGTLPGHMTERFQEAIDNLAVLMRAETGSRPYKAAAAQLPETLETIMLLGLLGTVSLGIFFVLMRNKGI
GKMGFGMVTLGASAWLMWLSEIEPARIACVLIVVFLLLVVLIPEPEKQRSPQDNQMAIIIMVAVGLLGLITANELGWLERTKSDLSHLM
GRREEGATIGFSMDIDLRPASAWAIYAALTTFITPAVQHAVTTSYNNYSLMAMATQAGVLFGMGKGMPFYAWDFGVPLMIGCYSQLT
PLTLIVAIILLVAHYMYLIPGLQAAAARAAQKRTAAGIMKNPVVDGIVVTDIDTMTIDPQVEKKMGQVLLIAVAVSSAILSRTAWGWGE
AGALITAATSTLWEGSPNKYWNSSTATSLCNIFRGSYLAGASLIYTVTRNAGLVKRRGGGTGETLGEKWKARLNQMSALEFYSYKKSG
ITEVCREEARRALKDGVATGGHAVSRGSAKLRWLVERGYLQPYGKVIDLGCGRGGWSYYAATIRKVQEVKGYTKGGPGHEEPMLVQ
SYGWNIVRLKSGVDVFHMAEPCDTLLCDIGESSSSPEVEEARTLRVLSMVGDWLEKRPGAFCIKVLCPYTSTMMETLERLQRRYGGGL
VRVPLSRNSTHEMYWVSGAKSNTIKSVSTTSQLLLGRMDGPRRPVKYEEDVNLGSGTRAVVSCAEAPNMKIIGNRIERIRSEHAETWFF
DENHPYRTWAYHGSYEAPTQGSASSLINGVVRLLSKPWDVVTGVTGIAMTDTTPYGQQRVFKEKVDTRVPDPQEGTRQVMSMVSSW
LWKELGKHKRPRVCTKEEFINKVRSNAALGAIFEEEKEWKTAVEAVNDPRFWALVDKEREHHLRGECQSCVYNMMGKREKKQGEFG
KAKGSRAIWYMWLGARFLEFEALGFLNEDHWMGRENSGGGVEGLGLQRLGYVLEEMSRIPGGRMYADDTAGWDTRISRFDLENEAL
ITNQMEKGHRALALAIIKYTYQNKVVKVLRPAEKGKTVMDIISRKDQRGSGQVVTYALNTFTNLVVQLIRNMEAEEVLEMQDLWLLR
RSEKVTNWLQSNGWDRLKRMAVSGDDCVVKPIDDRFAHALRFLNDMGKVRKDTQEWKPSTGWDNWEEVPFCSHHFNKLHLKDGRS
IVVPCRHQDELIGRARVSPGAGWSIRETACLAKSYAQMWQLLYFHRRDLRLMANAICSSVPVDWVPTGRTTWSIHGKGEWMTTEDML
VVWNRVWIEENDHMEDKTPVTKWTDIPYLGKREDLWCGSLIGHRPRTTWAENIKNTVNMVRRIIGDEEKYMDYLSTQVRYLGEEGST
PGVL

2. Zika virus genomic RNA, complete genome, strain: ZIKV/Hu/NIID123/2016

GenBank: LC219720.1

MKNPKKKSGGFRIVNMLKRGVARVSPFGGLKRLPAGLLLGHGPIRMVLAILAFLRFTAIKPSLGLINRWGSVGKKEAMEIIKKFKKDLA
AMLRIINARKEKKRRGADTSVGIVGLLLTTAMAAEVTRRGSAYYMYLDRSDAGEAISFPTTLGMNKCHIQIMDLGHMCDATMSYECP
MLDEGVEPDDVDCWCNTTSTWVVYGTCHHKKGEARRSRRAVTLPSHSTRKLQTRSQTWLESREYTKHLIRVENWIFRNPGFALAAAA

IAWLFGSSTSQKVIYLVMILLIAPAYSIRCIGVSNRDFVEGMSGGTWVDVVLEHGGCVTVMAQDKPTVDIELVTTTVSNMAEVRSYCYE
ASISDMASDSRCPTQGEAYLDRQSDTQYVCKRTLVDRGWGNGCGLFGKGSLVTCAKFACSKKMTGKSIQPENLEYRIMLSVHGSQHS
GMIVNDTGHETDENRAKVEITPNSPRAEATLGGFGSLGLDCEPRTGLDFSDLYYLTMNNKHWLVHKEWFHDIPLPWHAGADTGTPHW
NNKEALVEFKDAHAKRQTVVVLGNQEGAVHTALAGALEAEMDGAKGRLSSGHLKCRLKMDKLRLKGVSYSLCTAAFTFTKIPAETL
HGTVTVEVQYAGTDGPCKVPAQMAVDMQTLTPVGRLITANPVITESTENSKMMLELDPPFGDSYIVIGVGEKKITHHWHRSGSTIGKAF
EATVRGAKRMAVLGDTAWDFGSVGGALNSLGKGIHQIFGAAFKSLFGGMSWFSQILIGTLLMWLGLNTKNGSISLMCLALGGVLIFLS
TAVSADVGCSVDFSKKETRCGTGVFVYNDVEAWRDRYKYHPDSPRRLAAAVKQAWEDGICGISSVSRMENIMWRSVEGELNAILEEN
GVQLTVVVGSVKNPMWRGPQRLPAPVNELPHGWKAWGKSYFVRAAKTNNSFVVDGDTLKECPLKHRAWNSFLVEDHGFGVFHTSV
WLKVREDYSLECDPAVIGTAVKGKEAVHSDLGYWIESEKNDTWRLKRAHLIEMKTCEWPKSHTLWTDGIEESDLIIPKSLAGPLSHHNT
REGYRTQMKGPWHSEELEIRFEECPGTKVHVEETCGTGPSLRSTTASGRVIEEWCCRECTMPPLSFRAKDGCWYGMEIRPRKEPESNLV
RSMVTAGSTDHIDHFSLGVLVILLMVQEGLKKRMTTKIIISTSMAVLVAMILGGFSMSDLAKLAILMGATFAEMNTGGDVAHLALIAAF
KVRPALLVSFIFRANWTPRESMLLALASCLLQTAISALEGDLMVLINGFALAWLAIRAMVVPRTDNITLAILAALTPLARGTLLVAWRA
GLATCGGFMLLSLKGKGSVKKNLPFVMALGLTAVRLVDPINVVGLLLLTRSGKRSWPPSEVLTAVGLICALAGGFAKADIEMAGPMA
AVGLLIVSYVVSGKSVDMYIERAGDITWEKDAEVTGNSPRLDVALDESGDFSLVEDDGPPMREIILKVVLMTICGMNPIAIPFAAGAWY
VYVKTGKRSGALWDVPAPKEVKKGETTDGVYRVMTRRLLGSTQVGVGVMQEGVFHTMWHVTKGSALRSGEGRLDPYWGDVKQDL
VSYCGPWKLDAAWDGHSEVQLLAVPPGERARNIQTLPGIFKTKDGDIGAVALDYPAGTSGSPILDKCGRVIGLYGNGVVIKNGSYVSAI
TQGRREEETPVECFEPSMLKKKQLTVLDLHPGAGKTRRVLPEIVREAIKTRLRTVILAPTRVVAAEMEEALRGLPVRYMTTAVNVTHSG
TEIVDLMCHATFTSRLLQPIRVPNYNLYIMDEAHFTDPSSIAARGYISTRVEMGEAAAIFMTATPPGTRDAFPDSNSPIMDTEVEPERAWS
SGFDWVTDHSGKTVWFVPSVRNGNEIAACLTKAGKRVIQLSRKTFETEFQKTKHQEWDFVVTTDISEMGANFKADRVIDSRRCLKPVI
LDGERVILAGPMPVTHASAAQRRGRIGRNPNKPGDEYLYGGGCAETDEDHAHWLEARMLLDNIYLQDGLIASLYRPEADKVAAIEGEF
KLRTEQRKTFVELMKRGDLPVWLAYQVASAGITYTDRRWCFDGTTNNTIMEDSVPAEVWTRYGEKRVLKPRWMDARVCSDHAALK
SFKEFAAGKRGAAFGVMEALGTLPGHMTERFQEAIDNLAVLMRAETGSRPYKAAAAQLPETLETIMLLGLLGTVSLGIFFVLMRNKGM
GKMGFGMVTLGASAWLMWLSEIEPARIACVLIVVFLLLVVLIPEPEKQRSPQDNQMAIIIMVAVGLLGLITANELGWLERTKSDLSHLM
GRREERATIGFSMDIDLRPASAWAIYAALTTFITPAVQHAVTTSYNNYSLMAMATQAGVLFGMGKGMPFYAWDFGVPLMIGCYSQLT
PLTLIVAIILLVAHYMYLIPGLQAAAARAAQKRTAAGIMKNPVVDGIVVTDIDTMTIDPQVEKKMGQVLLIAVAVSSAILSRTAWGWGE
AGALITAATSTLWEGSPNKYWNSSTATSLCNIFRGSYLAGASLIYTVTRNAGLVKRRGGGTGETLGEKWKARLNQMSALEFYSYKKSG
ITEVCREEARRALKDGVATGGHAVSRGSAKLRWLVERGYLQPYGKVIDLGCGRGGWSYYAATIRKVQEVKGYTKGGPGHEEPMLVQ
SYGWNIVRLKSGVDVFHMAEPCDTLLCDIGESSSSPEVEEARTLRVLSMVGDWLEKRPGAFCIKVLCPYTSTMMETLERLQRRYGGGL
VRVPLSRNSTHEMYWVSGAKSNTIKSVSTTSQLLLGRMDGPRRPVKYEEDVNLGSGTRAVVSCAEAPNMKIIGNRIERIRSEHAETWFF
DENHPYRTWAYHGSYEAPTQGSASSLINGVVRLLSKPWDVVTGVTGIAMTDTTPYGQQRVFKEKVDTRVPDPQEGTRQVMSMVSSW
LWKELGKHKRPRVCTKEEFINKVRSNAALGAIFEEEKEWKTAVEAVNDPRFWALVDKEREHHLRGECQSCVYNMMGKREKKQGEFG
KAKGSRAIWYMWLGARFLEFEALGFLNEDHWMGRENSGGGVEGLGLQRLGYVLEEMSRIPGGRMYADDTAGWDTRISRFDLENEAL
ITNQMEKGHRALALAIIKYTYQNKVVKVLRPAEKGKTVMDIISRQDQRGSGQVVTYALNTFTNLVVQLIRNMEAEEVLEMQDLWLLR
RSEKVTNWLQSNGWDRLKRMAVSGDDCVVKPIDDRFAHALRFLNDMGKVRKDTQEWRPSTGWDNWEEVPFCSHHFNKLHLKDGRS
IVVPCRHQDELIGRARVSPGAGWSIRETACLAKSYAQMWQLLYFHRRDLRLMANAICSSVPVDWVPTGRTTWSIHGKGEWMTTEDML
VVWNRVWIEENDHMEDKTPVTKWTDIPYLGKREDLWCGSLIGHRPRTTWAENIKNTVNIVRRIIGDEEKYMDYLSTQVRYLGEEGSTP
GVL

3. Zika virus genomic RNA, complete genome, strain: ZIKV/Hu/Chiba/S36/2016

GenBank: LC191864.1

MKNPKKKSGGFRIVNMLKRGVARVSPFGGLKRLPAGLLLGHGPIRMVLAILAFLRFTAIKPSLGLINRWGSVGKKEAME
IIKKFKKDLAAMLRIINARKEKKRRGADTNVGIVGLLLTTAMAAEVTRRGSAYYMYLDRNDAGEAISFPTTLGMNKCYI
QIMDLGHMCDATMSYECPMLDEGVEPDDVDCWCNMTSTWVVYGTCHHKKGEARRSRRAVTLPSHSTRKLQTRSQT
WLESREYTKHLIRVENWIFRNPGFALAAAAIAWLLGSSTSQKVIYLVMILLIAPAYSIRCIGVSNRDFVEGMSGGTWVDV
VLEHGGCVTVMAQDKPTVDIELVTTTVSNMAEVRSYCYEASISDMASDSRCPTQGEAYLDKQSDTQYVCKRTLVDRG
WGNGCGLFGKGSLVTCAKFACSKKMTGKSIQPENLEYRIMLSVHGSQHSGMIVNDTGHETDENRAKVEITPNSPRAEA
TLGGFGSLGLDCEPRTGLDFSDLYYLTMNNKHWLVHKEWFHDIPLPWHAGADTGTPHWNNKEALVEFKDAHAKRQT
VVVLGSQEGAVHTALAGALEAEMDGAKGRLSSGHLKCRLKMDKLRLKGVSYSLCTAAFTFTKIPAETLHGTVTVEVQ
YAGTDGPCKVPAQMAVDMQTLTPVGRLITANPVITESTENSKMMLELDPPFGDSYIVIGVGEKKITHHWHRSGSTIGKA

FEATVRGAKRMAVLGDTAWDFGSVGGALNSLGKGIHQIFGAAFKSLFGGMSWFSQILIGTLLMWLGLNTKNGSISLMC
LALGGVLIFLSTAVSADVGCSVDFSKKETRCGTGVFVYNDVEAWRDRYKYHPDSPRRLAAAVKQAWEDGICGISSVSR
MENIMWRSVEGELNAILEENGVQLTVVVGSVKNPMWRGPQRLPVPVNELPHGWKAWGKSYFVRAAKTNNSFVVDG
DTLKECPLKHRAWNSFLVEDHGFGVFHTSVWLKVREDYSLECDSAVIGTAVKGKEAVHSDLGYWIESEKNDTWRLKR
AHLIEMKTCEWPKSHTLWTDGIEESDLIIPKSLAGPLSHHNTREGYRTQMKGPWHSEELEIRFEECPGTKVHVEETCGTG
PSLRSTTASGRVIEEWCCRECTMPPLSFRAKDGCWYGMEIRPRKEPESNLVRSMVTAGSTDHMDHFSLGVLVILLMVQE
GLKKRMTTKIIISTSMAVLVAMILGGFSMSDLAKLAILMGATFAEMNTGGDVAHLALIAAFKVRPALLVSFIFRANWTP
RESMLLALASCLLQTAISALEGDLMVLINGFALAWLAIRAMVVPRTDNITLAILAALTPLARGTLLVAWRAGLATCGGF
MLLSLKGKGSVKKNLPFVMALGLTAVRLVDPINVVGLLLLTRSGKRSWPPSEVLTAVGLICALAGGFAKADIEMAGPM
AAVGLLIVSYVVSGKSVDMYIERAGDITWEKDAEVTGNSPRLDVALDESGDFSLVEDDGPPMREIILKVVLMTICGMNP
IAIPFAAGAWYVYVKTGKRSGALWDVPAPKEVKKGETTDGVYRVMTRRLLGSTQVGVGVMQEGVFHTMWHVTKGS
ALRSGEGRLDPYWGDVKQDLVSYCGPWKLDAAWDGHSEVQLLAVPPGERARNIQTLPGIFKTKDGDIGAVALDYPAG
TSGSPILDKCGRVIGLYGNGVVIKNGSYVSAITQGRREEETPVECFEPSMLKKKQLTVLDLHPGAGKTRRVLPEIVREAIK
TRLRTVILAPTRVVAAEMEEALRGLPVRYMTTAVNVTHSGTEIVDLMCHATFTSRLLQPIRVPNYNLYIMDEAHFTDPS
SIAARGYISTRVEMGEAAAIFMTATPPGTRDAFPDSNSPIMDTEVEPERAWSSGFDWVTDHSGKTVWFVPSVRNGNEIA
ACLTKAGKRVIQLSRKTFETEFQKTKHQEWDFVVTTDISEMGANFKADRVIDSRRCLKPVILDGERVILAGPMPVTHAS
AAQRRGRIGRNPNKPGDEYLYGGGCAETDEDHAHWLEARMLLDNIYLQDGLIASLYRPEADKVAAIEGEFKLRTEQRK
TFVELMKRGDLPVWLAYQVASAGITYTDRRWCFDGTTNNTIMEDSVPAEVWTRHGEKRVLKPRWMDARVCSDHAAL
KSFKEFAAGKRGAAFGVMEALGTLPGHMTERFQEAIDNLAVLMRAETGSRPYKAAAAQLPETLETIMLLGLLGTVSLG
IFFVLMRNKGIGKMGFGMVTLGASAWLMWLSEIEPARIACVLIVVFLLLVVLIPEPEKQRSPQDNQMAIIIMVAVGLLGL
ITANELGWLERTKSDLSHLMGRREEGATIGFSMDIDLRPASAWAIYAALTTFITPAVQHAVTTSYNNYSLMAMATQAG
VLFGMGKGMPFYAWDFGVPLMIGCYSQLTPLTLIVAIILLVAHYMYLIPGLQAAAARAAQKRTAAGIMKNPVVDGIVV
TDIDTMTIDPQVEKKMGQVLLIAVAVSSAILSRTAWGWGEAGALITAATSTLWEGSPNKYWNSSTATSLCNIFRGSYLA
GASLIYTVTRNAGLVKRRGGGTGETLGEKWKARLNQMSALEFYSYKKSGITEVCREEARRALKDGVATGGHAVSRGS
AKLRWLVERGYLQPYGKVIDLGCGRGGWSYYAATIRKVQEVKGYTKGGPGHEEPMLVQSYGWNIVRLKSGVDVFHM
AEPCDTLLCDIGESSSSPEVEEARTLRVLSMVGDWLEKRPGAFCIKVLCPYTSTMMETLERLQRRYGGGLVRVPLSRNS
THEMYWVSGAKSNTIKSVSTTSQLLLGRMDGPRRPVKYEEDVNLGSGTRAVVSCAEAPNMKIIGNRIERIRSEHAETWF
FDENHPYRTWAYHGSYEAPTQGSASSLINGVVRLLSKPWDVVTGVTGIAMTDTTPYGQQRVFKEKVDTRVPDPQEGTR
QVMSMVSSWLWKELGKHKRPRVCTKEEFINKVRSSAALGAIFEEEKEWKTAVEAVNDPRFWALVDKEREHHLRGECQ
SCVYNMMGKREKKQGEFGKAKGSRAIWYMWLGARFLEFEALGFLNEDHWMGRENSGGGVEGLGLQRLGYVLEEMS
RIPGGRMYADDTAGWDTRISRFDLENEALITNQMEKGHRALALAIIKYTYHNKVVKVLRPAEKGKTVMDIISRQDQRG
SGQVVTYALNTFTNLVVQLIRSMEAEEVLEMQDLWLLRRSEKVTNWLQSNGWDRLKRMAVSGDDCVVKPIDDRFAH
ALRFLNDMGKVRKDTQEWKPSTGWDNWEEVPFCSHHFNKLHLKDGRSIVVPCRHQDELIGRARVSPGAGWSIRETAC
LAKSYAQMWQLLYFHRRDLRLMANAICSSVPVDWVPTGRTTWSIHGKGEWMTTEDMLVVWNRVWIEENDHMEDKT
PVTKWTDIPYLGKREDLWCGSLIGHRPRTTWAENIKNTVNMVRRIIGDEEKYMDYLSTQVRYLGEEGSTPGVL

4. Zika virus genomic RNA, complete genome, strain: ZIKV/Hu/Yokohama/1/2016

GenBank: LC190723.1

MKNPKKKSGGFRIVNMLKRGVARVNPLGGLKRLPAGLLLGHGPIRMVLAILAFLRFTAIKPSLGLINRWGSVGKKEAM
EIIKKFKKDLAAMLRIINARKERKRRGADTSIGIIGLLLTTAMAAEITRRGSAYYMYLDRSDAGKAISFATTLGVNKCHV
QIMDLGHMCDATMSYECPMLDEGVEPDDVDCWCNTTSTWVVYGTCHHKKGEARRSRRAVTLPSHSTRKLQTRSQTW
LESREYTKHLIKVENWIFRNPGFALVAVAIAWLLGSSTSQKVIYLVMILLIAPAYSIRCIGVSNRDFVEGMSGGTWVDVV
LEHGGCVTVMAQDKPTVDIELVTTTVSNMAEVRSYCYEASISDMASDSRCPTQGEAYLDKQSDTQYVCKRTLVDRGW
GNGCGLFGKGSLVTCAKFTCSKKMTGKSIQPENLEYRIMLSVHGSQHSGMTVNDIGYETDENRAKVEVTPNSPRAEAT
LGGFGSLGLDCEPRTGLDFSDLYYLTMNNKHWLVHKEWFHDIPLPWHAGADTGTPHWNNKEALVEFKDAHAKRQTV

VVLGSQEGAVHTALAGALEAEMDGAKGKLFSGHLKCRLKMDKLRLKGVSYSLCTAAFTFTKVPAETLHGTVTVEVQY
AGTDGPCKIPVQMAVDMQTLTPVGRLITANPVITESTENSKMMLELDPPFGDSYIVIGVGDKKITHHWHRSGSTIGKAFE
ATVRGAKRMAVLGDTAWDFGSVGGVFNSLGKGIHQIFGAAFKSLFGGMSWFSQILIGTLLVWLGLNTKNGSISLTCLA
LGGVMIFLSTAVSADVGCSVDFSKKETRCGTGVFIYNDVEAWRDRYKYHPDSPRRLAAAVKQAWEEGICGISSVSRME
NIMWKSVEGELNAILEENGVQLTVVVGSVKNPMWRGPQRLPVPVNELPHGWKAWGKSYFVRAAKTNNSFVVDGDTL
KECPLEHRAWNSFLVEDHGFGVFHTSVWLKVREDYSLECDPAVIGTAVKGREAAHSDLGYWIESEKNDTWRLKRAHL
IEMKTCEWPKSHTLWTDGVEESDLIIPKSLAGPLSHHNTREGYRTQVKGPWHSEELEIRFEECPGTKVYVEETCGTGPSL
RSTTASGRVIEEWCCRECTMPPLSFRAKDGCWYGMEIRPRKEPESNLVRSMVTAGSTDHMDHFSLGVLVILLMVQEGL
KKRMTTKIIMSTSMAVLVVMILGGFSMSDLAKLVILMGATFAEMNTGGDVAHLALVAAFKVRPALLVSFIFRANWTPR
ESMLLALASCLLQTAISALEGDLMVLINGFALAWLAIRAMAVPRTDNIALPILAALTPLARGTLLVAWRAGLATCGGIM
LLSLKGKGSVKKNLPFVMALGLTAVRVVDPINVVGLLLLTRSGKRSWPPSEVLTAVGLICALAGGFAKADIEMAGPMA
AVGLLIVSYVVSGKSVDMYIERAGDITWEKDAEVTGNSPRLDVALDESGDFSLVEEDGPPMREIILKVVLMAICGMNPI
AIPFAAGAWYVYVKTGKRSGALWDVPAPKEVKKGETTDGVYRVMTRRLLGSTQVGVGVMQEGVFHTMWHVTKGA
ALRSGEGRLDPYWGDVKQDLVSYCGPWKLDAAWDGLSEVQLLAVPPGERARNIQTLPGIFKTKDGDIGAVALDYPAG
TSGSPILDKCGRVIGLYGNGVVIKNGSYVSAITQGKREEETPVECFEPSMLKKKQLTVLDLHPGAGKTRRVLPEIVREAI
KKRLRTVILAPTRVVAAEMEEALRGLPVRYMTTAVNVTHSGTEIVDLMCHATFTSRLLQPIRVPNYNLYIMDEAHFTDP
SSIAARGYISTRVEMGEAAAIFMTATPPGTRDAFPDSNSPIMDTEVEPERAWSSGFDWVTDHSGKTVWFVPSVRNGNEI
AACLTKAGKRVIQLSRKTFETEFQKTKNQEWDFVITTDISEMGANFKADRVIDSRRCLKPVILDGERVILAGPMPVTHAS
AAQRRGRIGRNPNKPGDEYMYGGGCAETDEGHAHWLEARMLLDNIYLQDGLIASLYRPEADKVAAIEGEFKLRTEQR
KTFVELMKRGDLPVWLAYQVASAGITYTDRRWCFDGTTNNTIMEDSVPAEVWTKYGEKRVLKPRWMDARVCSDHA
ALKSFKEFAAGKRGAALGVMEALGTLPGHMTERFQEAIDNLAVLMRAETGSRPYKAAAAQLPETLETIMLLGLLGTVS
LGIFFVLMRNKGIGKMGFGMVTLGASAWLMWLSEIEPARIACVLIVVFLLLVVLIPEPEKQRSPQDNQMAIIIMVAVGLL
GLITANELGWLERTKNDIAHLMGRREEGATMGFSMDIDLRPASAWAIYAALTTLITPAVQHAVTTSYNNYSLMAMATQ
AGVLFGMGKGMPFYAWDLGVPLMMGCYSQLTPLTLIVAIILLVAHYMYLIPGLQAAAARAAQKRTAAGIMKNPVVDG
IVVTDIDTMTIDPQVEKKMGQVLLIAVAISSAVLLRTAWGWGEAGALITAATSTLWEGSPNKYWNSSTATSLCNIFRGS
YLAGASLIYTVTRNAGLVKRRGGGTGETLGEKWKARLNQMSALEFYSYKKSGITEVCREEARRALKDGVATGGHAVS
RGSAKLRWLVERGYLQPYGKVVDLGCGRGGWSYYAATIRKVQEVRGYTKGGPGHEEPMLVQSYGWNIVRLKSGVD
VFHMAEPCDTLLCDIGESSSSPEVEETRTLRVLSMVGDWLEKRPGAFCIKVLCPYTSTMMETMERLQRRHGGGLVRVP
LSRNSTHEMYWVSGAKSNIIKSVSTTSQLLLGRMDGPRRPVKYEEDVNLGSGTRAVASCAEAPNMKIIGRRIERIRNEHA
ETWFLDENHPYRTWAYHGSYEAPTQGSASSLVNGVVRLLSKPWDVVTGVTGIAMTDTTPYGQQRVFKEKVDTRVPDP
QEGTRQVMNIVSSWLWKELGKRKRPRVCTKEEFINKVRSNAALGAIFEEEKEWKTAVEAVNDPRFWALVDREREHHL
RGECHSCVYNMMGKREKKQGEFGKAKGSRAIWYMWLGARFLEFEALGFLNEDHWMGRENSGGGVEGLGLQRLGYI
LEEMNRAPGGKMYADDTAGWDTRISKFDLENEALITNQMEEGHRTLALAVIKYTYQNKVVKVLRPAEGGKTVMDIIS
RQDQRGSGQVVTYALNTFTNLVVQLIRNMEAEEVLEMQDLWLLRKPEKVTRWLQSNGWDRLKRMAVSGDDCVVKPI
DDRFAHALRFLNDMGKVRKDTQEWKPSTGWSNWEEVPFCSHHFNKLYLKDGRSIVVPCRHQDELIGRARVSPGAGWS
IRETACLAKSYAQMWQLLYFHRRDLRLMANAICSAVPVDWVPTGRTTWSIHGKGEWMTTEDMLMVWNRVWIEEND
HMEDKTPVTKWTDIPYLGKREDLWCGSLIGHRPRTTWAENIKDTVNMVRRIIGDEEKYMDYLSTQVRYLGEEGSTPGV
L

5. Zika virus genomic RNA, complete genome, strain: MR766-NIID

GenBank: LC002520.1

MKNPKKKSGGFRIVNMLKRGVARVSPFGGLKRLPAGLLLGHGPIRMVLAILAFLRFTAIKPSLGLINRWGSVGKKEAME
IIKKFKKDLAAMLRIINARKEKKRRGADTSVGIVGLLLTTAMAAEVTRRGSAYYMYLDRNDAGEAISFPTTLGMNKCYI
QIMDLGHMCDATMSYECPMLDEGVEPDDVDCWCNTTSTWVVYGTCHHKKGEARRSRRAVTLPSHSTRKLQTRSQTW
LESREYTKHLIRVENWIFRNPGFALAAAAIAWLLGSSTSQKVIYLVMILLIAPAYSIRCIGVSNRDFVEGMSGGTWVDVV

LEHGGCVTVMAQDKPTVDIELVTTTVSNMAEVRSYCYEASISDMASDSRCPTQGEAYLDKQSDTQYVCKRTLVDRGW
GNGCGLFGKGSLVTCAKFACSKKMTGKSIQPENLEYRIMLSVHGSQHSGMIVNDTGHETDENRAKVEITPNSPRAEATL
GGFGSLGLDCEPRTGLDFSDLYYLTMNNKHWLVHKEWFHDIPLPWHAGADTGTPHWNNKEALVEFKDAHAKRQTVV
VLGSQEGAVHTALAGALEAEMDGAKGRLSSGHLKCRLKMDKLRLKGVSYSLCTAAFTFTKIPAETLHGTVTVEVQYA
GTDGPCKVPAQMAVDMQTLTPVGRLITANPVITESAENSKMMLELDPPFGDSYIVIGVGEKKITHHWHRSGSTIGKAFE
ATVRGAKRMAVLGDTAWDFGSVGGALNSLGKGIHQIFGAAFKSLFGGMSWFSQVLIGTLLMWLGLNTKNGSISLMCL
ALGGVLIFLSTAVSADVGCSVDFSKKETRCGTGVFVYNDVEAWRDRYKYHPDSPRRLAAAVKQAWEDGICGISSVSR
MENIMWRSVEGELNAILEENGVQLTVVVGSVKNPMWRGPQRLPVPVNELPHGWKAWGKSYFVRAAKTNNSFVVDG
DTLKECPLKHRAWNSFLVEDHGFGVFHTSVWLKVREDYSLECDPAVIGTAVKGKEAVHSDLGYWIESEKNDTWRLKR
AHLIEMKTCEWPKSHTLWTDGIEESDLIIPKSLAGPLSHHNTREGYRTQMKGPWHSEELEIRFEECPGTKVHVEETCGTG
PSLRSTTASGRVIEEWCCRECTMPPLSFRAKDGCWYGMEIRPRKEPESNLVRSMVTAGSTDHMDHFSLGVLVILLMVQE
GLKKRMTTKIIISTSMAVLVAMILGGFSMSDLAKLAILMGATFAEMNTGGDVAHLALIAAFKVRPALLVSFIFRANWTP
RESMLLALASCLLQTAISALEGDLMVLINGFALAWLAIRAMVVPRTDNITLAILAALTPLARGTLLVAWRAGLATCGGF
MLLSLKGKGSVKKNLPFVMALGLTAVRLVDPINVVGLLLLTRSGKRSWPPSEVLTAVGLICALAGGFAKADIEMAGPM
AAVGLLIVSYVVSGKSVDMYIERAGDITWEKDAEVTGNSPRLDVALDESGDFSLVEDDGPPMREIILKVVLMTICGMNP
IAIPFAAGAWYVYVKTGKRSGALWDVPAPKEVKKGETTDGVYRVMTRRLLGSTQVGVGVMQEGVFHTMWHVTKGS
ALRSGEGRLDPYWGDVKQDLVSYCGPWKLDAAWDGHSEVQLLAVAPGERARNIQTLPGIFKTKDGDIGAVALDYPAG
TSGSPILDKCGRVIGLYGNGVVIKNGSYVSAITQGRREEETPVECFEPSMLKKKQLTVLDLHPGAGKTRRVLPEIVREAIK
TRLRTVILAPTRVVAAEMEEALRGLPVRYMTTAVNVTHSGTEIVDLMCHATFTSRLLQPIRVPNYNLYIMDEAHFTDPS
SIAARGYISTRVEMGEAAAIFMTATPPGTRDAFPDSNSPIMDTEVEPERAWSSGFDWVTDHSGKTVWFVPSVRNGNEIA
ACLTKAGKRVIQLSRKTFETEFQKTKHQEWDFVVTTDISEMGANFKADRVIDSRRCLKPVILDGERVILAGPMPVTHAS
AAQRRGRIGRNPNKPGDEYLYGGGCAETDEDHAHWLEARMLLDNIYLQDGLIASLYRPEADKVAAIEGEFKLRTEQRK
TFVELMKRGDLPVWLAYQVASAGITYTDRRWCFDGTTNNTIMEDSVPAEVWTRHGEKRVLKPRWMDARVCSDHAAL
KSFKEFAAGKRGAAFGVMEALGTLPGHMTERFQEAIDNLAVLMRAETGSRPYKAAAAQLPETLETIMLLGLLGTVSLG
IFFVLMRNKGIGKMGFGMVTLGASAWLMWLSEIEPARIACVLIVVFLLLVVLIPEPEKQRSPQDNQMAIIIMVAVGLLGL
ITANELGWLERTKSDLSHLMGRREEGATIGFSMDIDLRPASAWAIYAALTTFITPAVQHAVTTSYNNYSLMAMATQAG
VLFGMGKGMPFYAWDFGVPLMIGCYSQLTPLTLIVAIILLVAHYMYLIPGLQAAAARAAQKRTAAGIMKNPVVDGIVV
TDIDTMTIDPQVEKKMGQVLLIAVAVSSAILSRTAWGWGEAGALITAATSTLWEGSPNKYWNSSTATSLCNIFRGSYLA
GASLIYTVTRNAGLVKRRGGGTGETLGEKWKARLNQMSALEFYSYKKSGITEVCREEARRALKDGVATGGHAVSRGS
AKLRWLVERGYLQPYGKVIDLGCGRGGWSYYAATIRKVQEVKGYTKGGPGHEEPVLVQSYGWNIVRLKSGVDVFHM
AEPCDTLLCDIGESSSSPEVEEARTLRVLSMVGDWLEKRPGAFCIKVLCPYTSTMMETLERLQRRYGGGLVRVPLSRNS
THEMYWVSGAKSNTIKSVSTTSQLLLGRMDGPRRPVKYEEDVNLGSGTRAVVSCAEAPNMKIIGNRIERIRSEHAETWF
FDENHPYRTWAYHGSYEAPTQGSASSLINGVVRLLSKPWDVVTGVTGIAMTDTTPYGQQRVFKEKVDTRVPDPQEGTR
QVMSMVSSWLWKELGKHKRPRVCTKEEFINKVRSNAALGAIFEEEKEWKTAVEAVNDPRFWALVDKEREHHLRGEC
QSCVYNMMGKREKKQGEFGKAKGSRAIWYMWLGARFLEFEALGFLNEDHWMGRENSGGGVEGLGLQRLGYVLEEM
SRIPGGRMYADDTAGWDTRISRFDLENEALITNQMEKGHRALALAIIKYTYQNKVVKVLRPAEKGKTVMDIISRQDQR
GSGQVVTYALNTFTNLVVQLIRNMEAEEVLEMQDLWLLRRSEKVTNWLQSNGWDRLKRMAVSGDDCVVKPIDDRFA
HALRFLNDMGKVRKDTQEWKPSTGWDNWEEVPFCSHHFNKLHLKDGRSIVVPCRHQDELIGRARVSPGAGWSIRETA
CLAKSYAQMWQLLYFHRRDLRLMANAICSSVPVDWVPTGRTTWSIHGKGEWMTTEDMLVVWNRVWIEENDHMEDK
TPVTKWTDIPYLGKREDLWCGSLIGHRPRTTWAENIKNTVNMVRRIIGDEEKYMDYLSTQVRYLGEEGSTPGVL

6. Zika virus isolate ZIKV/Homo_sapiens/CPV/2016/CV907u, complete genome

GenBank: MK241417.1

MKNPKKKSGGFRIVNMLKRGVARVSPFGGLKRLPAGLLLGHGPIRMVLAILAFLRFTAIKPSLGLINRWGSVGKKEAME
IIKKFKKDLAAMLRIINARKEKKRRGADTSVGIVGLLLTTAMAAEVTRRGSAYYMYLDRNDAGEAISFPTTLGMNKCYI

QIMDLGHMCDATMSYECPMLDEGVEPDDVDCWCNTTSTWVVYGTCHHKKGEARRSRRAVTLPSHSTRKLQTRSQTW
LESREYTKHLIRVENWIFRNPGFALAAAAIAWLLGSSTSQKVIYLVMILLIAPAYSIRCIGVSNRDFVEGMSGGTWVDVV
LEHGGCVTVMAQDKPTVDIELVTTTVSNMAEVRSYCYEASISDMASDSRCPTQGEAYLDKQSDTQYVCKRTLVDRGW
GNGCGLFGKGSLVTCAKFACSKKMTGKSIQPENLEYRIMLSVHGSQHSGMIVNDTGHETDENRAKVEITPNSPRAEATL
GGFGSLGLDCEPRTGLDFSDLYYLTMNNKHWLVHKEWFHDIPLPWHAGADTGTPHWNNKEALVEFKDAHAKRQTVV
VLGSQEGAVHTALAGALEAEMDGAKGRLSSGHLKCRLKMDKLRLKGVSYSLCTAAFTFTKIPAETLHGTVTVEVQYA
GTDGPCKVPAQMAVDMQTLTPVGRLITANPVITESTENSKMMLELDPPFGDSYIVIGVGEKKITHHWHRSGSTIGKAFE
ATVRGAKRMAVLGDTAWDFGSVGGALNSLGKGIHQIFGAAFKSLFGGMSWFSQVLIGTLLMWLGLNTKNGSISLMCL
ALGGVLIFLSTAVSADVGCSVDFSKKETRCGTGVFVYNDVEAWRDRYKYHPDSPRRLAAAVKQAWEDGICGISSVSR
MENIMWRSVEGELNAILEENGVQLTVVVGSVKNPMWRGPQRLPVPVNELPHGWKAWGKSYFVRAAKTNNSFVVDG
DTLKECPLKHRAWNSFLVEDHGFGVFHTSVWLKVREDYSLECDPAVIGTAVKGKEAVHSDLGYWIESEKNDTWRLKR
AHLIEMKTCEWPKSHTLWTDGIEESDLIIPKSLAGPLSHHNTREGYRTQMKGPWHSEELEIRFEECPGTKVHVEETCGTG
PSLRSTTASGRVIEEWCCRECTMPPLSFRAKDGCWYGMEIRPRKEPESNLVRSMVTAGSTDHMDHFSLGVLVILLMVQE
GLKKRMTTKIIISTSMAVLVAMILGGFSMSDLAKLAILMGATFAEMNTGGDVAHLALIAAFKVRPALLVSFIFRANWTP
RESMLLALASCLLQTAISALEGDLMVLINGFALAWLAIRAMVVPRTDNITLAILAALTPLARGTLLVAWRAGLATCGGF
MLLSLKGKGSVKKNLPFVMALGLTAVRLVDPINVVGLLLLTRSGKRSWPPSEVLTAVGLICALAGGFAKADIEMAGPM
AAVGLLIVSYVVSGKSVDMYIERAGDITWEKDAEVTGNSPRLDVALDESGDFSLVEDDGPPMREIILKVVLMTICGMNP
IAIPFAAGAWYVYVKTGKRSGALWDVPAPKEVKKGETTDGVYRVMTRRLLGSTQVGVGVMQEGVFHTMWHVTKGS
ALRSGEGRLDPYWGDVKQDLVSYCGPWKLDAAWDGHSEVQLLAVPPGERARNIQTLPGIFKTKDGDIGAVALDYPAG
TSGSPILDKCGRVIGLYGNGVVIKNGSYVSAITQGRREEETPVECFEPSMLKKKQLTVLDLHPGAGKTRRVLPEIVREAIK
TRLRTVILAPTRVVAAEMEEALRGLPVRYMTTAVNVTHSGTEIVDLMCHATFTSRLLQPIRVPNYNLYIMDEAHFTDPS
SIAARGYISTRVEMGEAAAIFMTATPPGTRDAFPDSNSPIMDTEVEPERAWSSGFDWVTDHSGKTVWFVPSVRNGNEIA
ACLTKAGKRVIQLSRKTFETEFQKTKHQEWDFVVTTDISEMGANFKADRVIDSRRCLKPVILDGERVILAGPMPVTHAS
AAQRRGRIGRNPNKPGDEYLYGGGCAETDEDHAHWLEARMLLDNIYLQDGLIASLYRPEADKVAAIEGEFKLRTEQRK
TFVELMKRGDLPVWLAYQVASAGITYTDRRWCFDGTTNNTIMEDSVPAEVWTRHGEKRVLKPRWMDARVCSDHAAL
KSFKEFAAGKRGAAFGVMEALGTLPGHMTERFQEAIDNLAVLMRAETGSRPYKAAAAQLPETLETIMLLGLLGTVSLG
IFFVLMRNKGIGKMGFGMVTLGASAWLMWLSEIEPARIACVLIVVFLLLVVLIPEPEKQRSPQDNQMAIIIMVAVGLLGL
ITANELGWLERTKSDLSHLMGRREEGATIGFSMDIDLRPASAWAIYAALTTFITPAVQHAVTTSYNNYSLMAMATQAG
VLFGMGKGMPFYAWDFGVPLMIGCYSQLTPLTLIVAIILLVAHYMYLIPGLQAAAARAAQKRTAAGIMKNPVVDGIVV
TDIDTMTIDPQVEKKMGQVLLIAVAVSSAILSRTAWGWGEAGALITAATSTLWEGSPNKYWNSSTATSLCNIFRGSYLA
GASLIYTVTRNAGLVKRRGGGTGETLGEKWKARLNQMSALEFYSYKKSGITEVCREEARRALKDGVATGGHAVSRGS
AKLRWLVERGYLQPYGKVIDLGCGRGGWSYYAATIRKVQEVKGYTKGGPGHEEPVLVQSYGWNIVRLKSGVDVFHM
AEPCDTLLCDIGESSSSPEVEEARTLRVLSMVGDWLEKRPGAFCIKVLCPYTSTMMETLERLQRRYGGGLVRVPLSRNS
THEMYWVSGAKSNTIKSVSTTSQLLLGRMDGPRRPVKYEEDVNLGSGTRAVVSCAEAPNMKIIGNRIERIRSEHAETWF
FDENHPYRTWAYHGSYEAPTQGSASSLINGVVRLLSKPWDVVTGVTGIAMTDTTPYGQQRVFKEKVDTRVPDPQEGTR
QVMSMVSSWLWKELGKHKRPRVCTKEEFINKVRSNAALGAIFEEEKEWKTAVEAVNDPRFWALVDKEREHHLRGEC
QSCVYNMMGKREKKQGEFGKAKGSRAIWYMWLGARFLEFEALGFLNEDHWMGRENSGGGVEGLGLQRLGYVLEEM
SRIPGGRMYADDTAGWDTRISRFDLENEALITNQMEKGHRALALAIIKYTYQNKVVKVLRPAEKGKTVMDIISRQDQR
GSGQVVTYALNTFTNLVVQLIRNMEAEEVLEMQDLWLLRRSEKVTNWLQSNGWDRLKRMAVSGDDCVVKPIDDRFA
HALRFLNDMGKVRKDTQEWKPSTGWDNWEEVPFCSHHFNKLHLKDGRSIVVPCRHQDELIGRARVSPGAGWSIRETA
CLAKSYAQMWQLLYFHRRDLRLMANAICSSVPVDWVPTGRTTWSIHGKGEWMTTEDMLVVWNRVWIEENDHMEDK
TPVTKWTDIPYLGKREDLWCGSLIGHRPRTTWAENIKNTVNMVRRIIGDEEKYMDYLSTQVRYLGEEGSTPGVL

7. Zika virus isolate ZIKV/Homo_sapiens/CPV/2015/CV487, complete genome

GenBank: MK241416.1

MKNPKKKSGGFRIVNMLKRGVARVSPFGGLKRLPAGLLLGHGPIRMVLAILAFLRFTAIKPSLGLINRWGSVGKKEAMEIIKKFKKDLA
AMLRIINARKEKKRRGADTSVGIVGLLLTTAMAAEVTRRGSAYYMYLDRNDAGEAISFPTTLGMNKCYIQIMDLGHMCDATMSYECP
MLDEGVEPDDVDCWCNTTSTWVVYGTCHHKKGEARRSRRAVTLPSHSTRKLQTRSQTWLESREYTKHLIRVENWIFRNPGFALAAAA
IAWLLGSSTSQKVIYLVMILLIAPAYSIRCIGVSNRDFVEGMSGGTWVDVVLEHGGCVTVMAQDKPTVDIELVTTTVSNMAEVRSYCYE
ASISDMASDSRCPTQGEAYLDKQSDTQYVCKRTLVDRGWGNGCGLFGKGSLVTCAKFACSKKMTGKSIQPENLEYRIMLSVHGSQHS
GMIVNDTGHETDENRAKVEITPNSPRAEATLGGFGSLGLDCEPRTGLDFSDLYYLTMNNKHWLVHKEWFHDIPLPWHAGADTGTPHW
NNKEALVEFKDAHAKRQTVVVLGSQEGAVHTALAGALEAEMDGAKGRLSSGHLKCRLKMDKLRLKGVSYSLCTAAFTFTKIPAETLH
GTVTVEVQYAGTDGPCKVPAQMAVDMQTLTPVGRLITANPVITESAENSKMMLELDPPFGDSYIVIGVGEKKITHHWHRSGSTIGKAFE
ATVRGAKRMAVLGDTAWDFGSVGGALNSLGKGIHQIFGAAFKSLFGGMSWFSQVLIGTLLMWLGLNTKNGSISLMCLALGGVLIFLST
AVSADVGCSVDFSKKETRCGTGVFVYNDVEAWRDRYKYHPDSPRRLAAAVKQAWEDGICGISSVSRMENIMWRSVEGELNAILEENG
VQLTVVVGSVKNPMWRGPQRLPVPVNELPHGWKAWGKSYFVRAAKTNNSFVVDGDTLKECPLKHRAWNSFLVEDHGFGVFHTSVW
LKVREDYSLECDPAVIGTAVKGKEAVHSDLGYWIESEKNDTWRLKRAHLIEMKTCEWPKSHTLWTDGIEESDLIIPKSLAGPLSHHNTR
EGYRTQMKGPWHSEELEIRFEECPGTKVHVEETCGTGPSLRSTTASGRVIEEWCCRECTMPPLSFRAKDGCWYGMEIRPRKEPESNLVR
SMVTAGSTDHMDHFSLGVLVILLMVQEGLKKRMTTKIIISTSMAVLVAMILGGFSMSDLAKLAILMGATFAEMNTGGDVAHLALIAAF
KVRPALLVSFIFRANWTPRESMLLALASCLLQTAISALEGDLMVLINGFALAWLAIRAMVVPRTDNITLAILAALTPLARGTLLVAWRA
GLATCGGFMLLSLKGKGSVKKNLPFVMALGLTAVRLVDPINVVGLLLLTRSGKRSWPPSEVLTAVGLICALAGGFAKADIEMAGPMA
AVGLLIVSYVVSGKSVDMYIERAGDITWEKDAEVTGNSPRLDVALDESGDFSLVEDDGPPMREIILKVVLMTICGMNPIAIPFAAGAWY
VYVKTGKRSGALWDVPAPREVKKGETTDGVYRVMTRRLLGSTQVGVGVMQEGVFHTMWHVTKGSALRSGEGRLDPYWGDVKQDL
VSYCGPWKLDAAWDGHSEVQLLAVPPGERARNIQTLPGIFKTKDGDIGAVALDYPAGTSGSPILDKCGRVIGLYGNGVVIKNGSYVSAI
TQGRREEETPVECFEPSMLKKKQLTVLDLHPGAGKTRRVLPEIVREAIKTRLRTVILAPTRVVAAEMEEALRGLPVRYMTTAVNVTHSG
TEIVDLMCHATFTSRLLQPIRVPNYNLYIMDEAHFTDPSSIAARGYISTRVEMGEAAAIFMTATPPGTRDAFPDSNSPIMDTEVEPERAWS
SGFDWVTDHSGKTVWFVPSVRNGNEIAACLTKAGKRVIQLSRKTFETEFQKTKHQEWDFVVTTDISEMGANFKADRVIDSRRCLKPVI
LDGERVILAGPMPVTHASAAQRRGRIGRNPNKPGDEYLYGGGCAETDEDHAHWLEARMLLDNIYLQDGLIASLYRPEADKVAAIEGEF
KLRTEQRKTFVELMKRGDLPVWLAYQVASAGITYTDRRWCFDGTTNNTIMEDSVPAEVWTRHGEKRVLKPRWMDARVCSDHAALK
SFKEFAAGKRGAAFGVMEALGTLPGHMTERFQEAIDNLAVLMRAETGSRPYKAAAAQLPETLETIMLLGLLGTVSLGIFFVLMRNKGI
GKMGFGMVTLGASAWLMWLSEIEPARIACVLIVVFLLLVVLIPEPEKQRSPQDNQMAIIIMVAVGLLGLITANELGWLERTKSDLSHLM
GRREEGATIGFSMDIDLRPASAWAIYAALTTFITPAVQHAVTTSYNNYSLMAMATQAGVLFGMGKGMPFYAWDFGVPLMIGCYSQLT
PLTLIVAIILLVAHYMYLIPGLQAAAARAAQKRTAAGIMKNPVVDGIVVTDIDTMTIDPQVEKKMGQVLLIAVAVSSAILSRTAWGWGE
AGALITAATSTLWEGSPNKYWNSSTATSLCNIFRGSYLAGASLIYTVTRNAGLVKRRGGGTGETLGEKWKARLNQMSALEFYSYKKSG
ITEVCREEARRALKDGVATGGHAVSRGSAKLRWLVERGYLQPYGKVIDLGCGRGGWSYYAATIRKVQEVKGYTKGGPGHEEPVLVQ
SYGWNIVRLKSGVDVFHMAEPCDTLLCDIGESSSSPEVEEARTLRVLSMVGDWLEKRPGAFCIKVLCPYTSTMMETLERLQRRYGGGL
VRVPLSRNSTHEMYWVSGAKSNTIKSVSTTSQLLLGRMDGPRRPVKYEEDVNLGSGTRAVVSCAEAPNMKIIGNRIERIRSEHAETWFF
DENHPYRTWAYHGSYEAPTQGSASSLINGVVRLLSKPWDVVTGVTGIAMTDTTPYGQQRVFKEKVDTRVPDPQEGTRQVMSMVSSW
LWKELGKHKRPRVCTKEEFINKVRSNAALGAIFEEEKEWKTAVEAVNDPRFWALVDKEREHHLRGECQSCVYNMMGKREKKQGEFG
KAKGSRAIWYMWLGARFLEFEALGFLNEDHWMGRENSGGGVEGLGLQRLGYVLEEMSRIPGGRMYADDTAGWDTRISRFDLENEAL
ITNQMEKGHRALALAIIKYTYQNKVVKVLRPAEKGKTVMDIISRQDQRGSGQVVTYALNTFTNLVVQLIRNMEAEEVLEMQDLWLLR
RSEKVTNWLQSNGWDRLKRMAVSGDDCVVKPIDDRFAHALRFLNDMGKVRKDTQEWKPSTGWDNWEEVPFCSHHFNKLHLKDGRS
IVVPCRHQDELIGRARVSPGAGWSIRETACLAKSYAQMWQLLYFHRRDLRLMANAICSSVPVDWVPTGRTTWSIHGKGEWMTTEDML
VVWNRVWIEENDHMEDKTPVTKWTDIPYLGKREDLWCGSLIGHRPRTTWAENIKNTVNMVRRIIGDEEKYMDYLSTQVRYLGEEGST
PGVL

8. Zika virus isolate ZIKV/Homo_sapiens/CPV/2015/CV448c, complete genome

GenBank: MK241415.1

MKNPKEEIRRIRIVNMLKRGVARVNPLGGLKRLPAGLLLGHGPIRMVLAILAFLRFTAIKPSLGLINRWGSVGKKEAMEIIKKFKKDLAA
MLRIINARKERKRRGADTSIGIIGLLLTTAMAAEITRRGSAYYMYLDRSDAGKAISFATTLGVNKCHVQIMDLGHMCDATMSYECPML
DEGVEPDDVDCWCNTTSTWVVYGTCHHKKGEARRSRRAVTLPSHSTRKLQTRSQTWLESREYTKHLIKVENWIFRNPGFALVAVAIA
WLLGSSTSQKVIYLVMILLIAPAYSIRCIGVSNRDFVEGMSGGTWVDVVLEHGGCVTVMAQDKPTVDIELVTTTVSNMAEVRSYCYEA
SISDMASDSRCPTQGEAYLDKQSDTQYVCKRTLVDRGWGNGCGLFGKGSLVTCAKFTCSKKMTGKSIQPENLEYRIMLSVHGSQHSG

MIGYETDEDRAKVEVTPNSPRAEATLGGFGSLGLDCEPRTGLDFSDLYYLTMNNKHWLVHKEWFHDIPLPWHAGADTGTPHWNNKE
ALVEFKDAHAKRQTVVVLGSQEGAVHTALAGALEAEMDGAKGRLFSGHLKCRLKMDKLRLKGVSYSLCTAAFTFTKVPAETLHGTV
TVEVQYAGTDGPCKIPVQMAVDMQTLTPVGRLITANPVITESTENSKMMLELDPPFGDSYIVIGVGDKKITHHWHRSGSTIGKAFEATV
RGAKRMAVLGDTAWDFGSVGGVFNSLGKGIHQIFGAAFKSLFGGMSWFSQILIGTLLVWLGLNTKNGSISLTCLALGGVMIFLSTAVSA
DVGCSVDFSKRETRCGTGVFIYNDVEAWRDRYKYHPDSPRRLAAAVKQAWEEGICGISSVSRMENIMWKSVEGELNAILEENGVQLTV
VVGSVKNPMWRGPQRLPVPVNELPHGWKAWGKSYFVRAAKTNNSFVVDGDTLKECPLEHRAWNSFLVEDHGFGVFHTSVWLKVRE
DYSLECDPAVIGTAVKGREAAHSDLGYWIESEKNDTWRLKRAHLIEMKTCEWPKSHTLWTDGVEESDLIIPKSLAGPLSHHNTREGYR
TQVKGPWHSEELEIRFEECPGTKVYVEETCGTGPSLRSTTASGRVIEEWCCRECTMPPLSFRAKDGCWYGMEIRPRKEPESNLVRSMVT
AGSTDHMDHFSLGVLVILLMVQEGLKKRMTTKIIMSTSMAVLVVMILGGFSMSDLAKLVILMGATFAEMNTGGDVAHLALVAAFKVR
PALLVSFIFRANWTPRESMLLALASCLLQTAISALEGDLMVLINGFALAWLAIRAMAVPRTDNIALPILAALTPLARGTLLVAWRAGLAT
CGGIMLLSLKGKGSVKKNLPFVMALGLTAVRVVDPINVVGLLLLTRSGKRSWPPSEVLTAVGLICALAGGFAKADIEMAGPMAAVGLL
IVSYVVSGKSVDMYIERAGDITWEKDAEVTGNSPRLDVALDESGDFSLVEEDGPPMREIILKVVLMAICGMNPIAIPFAAGAWYVYVKT
GKRSGALWDVPAPKEVKKGETTDGVYRVMTRRLLGSTQVGVGVMQEGVFHTMWHVTKGAALRSGEGRLDPYWGDVKQDLVSYCG
PWKLDAAWDGLSEVQLLAVPPGERARNIQTLPGIFKTKDGDIGAVALDYPAGTSGSPILDKCGRVIGLYGNGVVIKNGSYVSAITQGKR
EEETPVECFEPSMLKKKQLTVLDLHPGAGKTRRVLPEIVREAIKKRLRTVILAPTRVVAAEMEEALRGLPVRYMTTAVNVTHSGTEIVD
LMCHATFTSRLLQPIRVPNYNLNIMDEAHFTDPSSIAARGYISTRVEMGEAAAIFMTATPPGTRDAFPDSNSPIMDTEVEVPERAWSSGF
DWVTDHSGKTVWFVPSVRNGNEIAACLTKAGKRVIQLSRKTFETEFQKTKNQEWDFVITTDISEMGANFKADRVIDSRRCLKPVILDGE
RVILAGPMPVTHASAAQRRGRIGRNPNKPGDEYMYGGGCAETDEGHAHWLEARMLLDNIYLQDGLIASLYRPEADKVAAIEGEFKLR
TEQRKTFVELMKRGDLPVWLAYQVASAGITYTDRRWCFDGTTNNTIMEDSVPAEVWTKYGEKRVLKPRWMDARVCSDHAALKSFKE
FAAGKRGAALGVMEALGTLPGHMTERFQEAIDNLAVLMRAETGSRPYKAAAAQLPETLETIMLLGLLGTVSLGIFFVLMRNKGIGKMG
FGMVTLGASAWLMWLSEIEPARIACVLIVVFLLLVVLIPEPEKQRSPQDNQMAIIIMVAVGLLGLITANELGWLERTKNDIAHLMGRREE
GATMGFSMDIDLRPASAWAIYAALTTLITPAVQHAVTTSYNNYSLMAMATQAGVLFGMGKGMPFMHGDLGVPLLMMGCYSQLTPLT
LIVAIILLVAHYMYLIPGLQAAAARAAQKRTAAGIMKNPVVDGIVVTDIDTMTIDPQVEKKMGQVLLIAVAISSAVLLRTAWGWGEAG
ALITAATSTLWEGSPNKYWNSSTATSLCNIFRGSYLAGASLIYTVTRNAGLVKRRGGGTGETLGEKWKARLNQMSALEFYSYKKSGITE
VCREEARRALKDGVATGGHAVSRGSAKIRWLEERGYLQPYGKVVDLGCGRGGWSYYAATIRKVQEVRGYTKGGPGHEEPMLVQSY
GWNIVRLKSGVDVFHMAAEPCDTLLCDIGESSSSPEVEETRTLRVLSMVGDWLEKRPGAFCIKVLCPYTSTMMETMERLQRRHGGGLV
RVPLCRNSTHEMYWVSGAKSNIIKSVSTTSQLLLGRMDGPRRPVKYEEDVNLGSGTRAVASCAEAPNMKIIGRRIERIRNEHAETWFLD
ENHPYRTWAYHGSYEAPTQGSASSLVNGVVRLLSKPWDVVTGVTGIAMTDTTPYGQQRVFKEKVDTRVPDPQEGTRQVMNIVSSWL
WKELGKRKRPRVCTKEEFINKVRSNAALGAIFEEEKEWKTAVEAVNDPRFWALVDRREREHHLRGECHSCVYNMMGKREKKQGEFGK
AKGSRAIWYMWLGARFLEFEALGFLNEDHWMGRENSGGGVEGLGLQRLGYILEEMNRAPGGKMYADDTAGWDTRISKFDLENEALI
TNQMEEGHRTLALAVIKYTYQNKVVKVLRPAEGGKTVMDIISRQDQRGSGQVVTYALNTFTNLVVQLIRNMEAEEVLEMQDLWLLR
KPEKVTRWLQSNGWDRLKRMAVSGDDCVVKPIDDRFAHALRFLNDMGKVRKDTQEWKPSTGWSNWEEVPFCSHHFNKLYLKDGRS
IVVPCRHQDELIGRARVSPGAGWSIRETACLAKSYAQMWQLLYFHRRDLRLMANAICSAVPVDWVPTGRTTWSIHGKGEWMTTEDM
LMVWNRVWIEENDHMEDKTPVTKWTDIPYLGKREDLWCGSLIGHRPRTTWAENIKDTVNMVRRIIGDEEKYMDYLSTQVRYLGEEGS
TPGVL

9. Zika virus, complete genomeNCBI Reference Sequence: NC_012532.1

MKNPKEEIRRIRIVNMLKRGVARVNPLGGLKRLPAGLLLGHGPIRMVLAILAFLRFTAIKPSLGLINRWGSVGKKEAMEIIKKFKKDLAA
MLRIINARKERKRRGADTSIGIIGLLLTTAMAAEITRRGSAYYMYLDRSDAGKAISFATTLGVNKCHVQIMDLGHMCDATMSYECPML
DEGVEPDDVDCWCNTTSTWVVYGTCHHKKGEARRSRRAVTLPSHSTRKLQTRSQTWLESREYTKHLIKVENWIFRNPGFALVAVAIA
WLLGSSTSQKVIYLVMILLIAPAYSIRCIGVSNRDFVEGMSGGTWVDVVLEHGGCVTVMAQDKPTVDIELVTTTVSNMAEVRSYCYEA
SISDMASDSRCPTQGEAYLDKQSDTQYVCKRTLVDRGWGNGCGLFGKGSLVTCAKFTCSKKMTGKSIQPENLEYRIMLSVHGSQHSG
MIGYETDEDRAKVEVTPNSPRAEATLGGFGSLGLDCEPRTGLDFSDLYYLTMNNKHWLVHKEWFHDIPLPWHAGADTGTPHWNNKE
ALVEFKDAHAKRQTVVVLGSQEGAVHTALAGALEAEMDGAKGRLFSGHLKCRLKMDKLRLKGVSYSLCTAAFTFTKVPAETLHGTV
TVEVQYAGTDGPCKIPVQMAVDMQTLTPVGRLITANPVITESTENSKMMLELDPPFGDSYIVIGVGDKKITHHWHRSGSTIGKAFEATV
RGAKRMAVLGDTAWDFGSVGGVFNSLGKGIHQIFGAAFKSLFGGMSWFSQILIGTLLVWLGLNTKNGSISLTCLALGGVMIFLSTAVSA
DVGCSVDFSKRETRCGTGVFIYNDVEAWRDRYKYHPDSPRRLAAAVKQAWEEGICGISSVSRMENIMWKSVEGELNAILEENGVQLTV
VVGSVKNPMWRGPQRLPVPVNELPHGWKAWGKSYFVRAAKTNNSFVVDGDTLKECPLEHRAWNSFLVEDHGFGVFHTSVWLKVRE
DYSLECDPAVIGTAVKGREAAHSDLGYWIESEKNDTWRLKRAHLIEMKTCEWPKSHTLWTDGVEESDLIIPKSLAGPLSHHNTREGYR
TQVKGPWHSEELEIRFEECPGTKVYVEETCGTGPSLRSTTASGRVIEEWCCRECTMPPLSFRAKDGCWYGMEIRPRKEPESNLVRSMVT
AGSTDHMDHFSLGVLVILLMVQEGLKKRMTTKIIMSTSMAVLVVMILGGFSMSDLAKLVILMGATFAEMNTGGDVAHLALVAAFKVR

PALLVSFIFRANWTPRESMLLALASCLLQTAISALEGDLMVLINGFALAWLAIRAMAVPRTDNIALPILAALTPLARGTLLVAWRAGLAT
CGGIMLLSLKGKGSVKKNLPFVMALGLTAVRVVDPINVVGLLLLTRSGKRSWPPSEVLTAVGLICALAGGFAKADIEMAGPMAAVGLL
IVSYVVSGKSVDMYIERAGDITWEKDAEVTGNSPRLDVALDESGDFSLVEEDGPPMREIILKVVLMAICGMNPIAIPFAAGAWYVYVKT
GKRSGALWDVPAPKEVKKGETTDGVYRVMTRRLLGSTQVGVGVMQEGVFHTMWHVTKGAALRSGEGRLDPYWGDVKQDLVSYCG
PWKLDAAWDGLSEVQLLAVPPGERARNIQTLPGIFKTKDGDIGAVALDYPAGTSGSPILDKCGRVIGLYGNGVVIKNGSYVSAITQGKR
EEETPVECFEPSMLKKKQLTVLDLHPGAGKTRRVLPEIVREAIKKRLRTVILAPTRVVAAEMEEALRGLPVRYMTTAVNVTHSGTEIVD
LMCHATFTSRLLQPIRVPNYNLNIMDEAHFTDPSSIAARGYISTRVEMGEAAAIFMTATPPGTRDAFPDSNSPIMDTEVEVPERAWSSGF
DWVTDHSGKTVWFVPSVRNGNEIAACLTKAGKRVIQLSRKTFETEFQKTKNQEWDFVITTDISEMGANFKADRVIDSRRCLKPVILDGE
RVILAGPMPVTHASAAQRRGRIGRNPNKPGDEYMYGGGCAETDEGHAHWLEARMLLDNIYLQDGLIASLYRPEADKVAAIEGEFKLR
TEQRKTFVELMKRGDLPVWLAYQVASAGITYTDRRWCFDGTTNNTIMEDSVPAEVWTKYGEKRVLKPRWMDARVCSDHAALKSFKE
FAAGKRGAALGVMEALGTLPGHMTERFQEAIDNLAVLMRAETGSRPYKAAAAQLPETLETIMLLGLLGTVSLGIFFVLMRNKGIGKMG
FGMVTLGASAWLMWLSEIEPARIACVLIVVFLLLVVLIPEPEKQRSPQDNQMAIIIMVAVGLLGLITANELGWLERTKNDIAHLMGRREE
GATMGFSMDIDLRPASAWAIYAALTTLITPAVQHAVTTSYNNYSLMAMATQAGVLFGMGKGMPFMHGDLGVPLLMMGCYSQLTPLT
LIVAIILLVAHYMYLIPGLQAAAARAAQKRTAAGIMKNPVVDGIVVTDIDTMTIDPQVEKKMGQVLLIAVAISSAVLLRTAWGWGEAG
ALITAATSTLWEGSPNKYWNSSTATSLCNIFRGSYLAGASLIYTVTRNAGLVKRRGGGTGETLGEKWKARLNQMSALEFYSYKKSGITE
VCREEARRALKDGVATGGHAVSRGSAKIRWLEERGYLQPYGKVVDLGCGRGGWSYYAATIRKVQEVRGYTKGGPGHEEPMLVQSY
GWNIVRLKSGVDVFHMAAEPCDTLLCDIGESSSSPEVEETRTLRVLSMVGDWLEKRPGAFCIKVLCPYTSTMMETMERLQRRHGGGLV
RVPLCRNSTHEMYWVSGAKSNIIKSVSTTSQLLLGRMDGPRRPVKYEEDVNLGSGTRAVASCAEAPNMKIIGRRIERIRNEHAETWFLD
ENHPYRTWAYHGSYEAPTQGSASSLVNGVVRLLSKPWDVVTGVTGIAMTDTTPYGQQRVFKEKVDTRVPDPQEGTRQVMNIVSSWL
WKELGKRKRPRVCTKEEFINKVRSNAALGAIFEEEKEWKTAVEAVNDPRFWALVDREREHHLRGECHSCVYNMMGKREKKQGEFGK
AKGSRAIWYMWLGARFLEFEALGFLNEDHWMGRENSGGGVEGLGLQRLGYILEEMNRAPGGKMYADDTAGWDTRISKFDLENEALI
TNQMEEGHRTLALAVIKYTYQNKVVKVLRPAEGGKTVMDIISRQDQRGSGQVVTYALNTFTNLVVQLIRNMEAEEVLEMQDLWLLR
KPEKVTRWLQSNGWDRLKRMAVSGDDCVVKPIDDRFAHALRFLNDMGKVRKDTQEWKPSTGWSNWEEVPFCSHHFNKLYLKDGRS
IVVPCRHQDELIGRARVSPGAGWSIRETACLAKSYAQMWQLLYFHRRDLRLMANAICSAVPVDWVPTGRTTWSIHGKGEWMTTEDM
LMVWNRVWIEENDHMEDKTPVTKWTDIPYLGKREDLWCGSLIGHRPRTTWAENIKDTVNMVRRIIGDEEKYMDYLSTQVRYLGEEGS
TPGVL